Title:
Vocabulary Size and Reading Comprehension in Elementary Level Emirati Learners of English

Name: Laurence Kinsella

Vocabulary Size and Reading Comprehension in Elementary Level Emirati Learners of English

by

Laurence Kinsella

Submitted in Partial Fulfilment for the Degree of

Doctor of Philosophy

University of Bedfordshire

September, 2018

DECLARATION

I, Laurence Kinsella, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have drawn on or cited the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis or any part of it is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- None of this work has been published before submission,

ABSTRACT

The overall aim of this mixed methods study based on a sequential explanatory design was to provide new knowledge and understanding regarding vocabulary learning and reading comprehension among elementary level Emirati learners of English.

The low vocabulary sizes and poor reading performances of these learners are well documented (Davidson, Atkinson & Spring, 2011; O'Sullivan, 2009). It is also widely accepted that students with low vocabulary size are will not read efficiently (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt, Jiang & Grabe, 2011). However, there is still considerable debate on how best low level students might quickly develop their vocabulary and how any increase in vocabulary size impacts on reading comprehension skills (Schmitt, 2010b). Further, much of the research carried out in this area has been in the context of cross sectional studies in experimental conditions rather than in classrooms (Nation & Webb, 2011). The present study aimed to address these gaps through a longitudinal classroom based study on the effect of word cards on receptive vocabulary size development. The quantitative experimental element of the design included an intervention using word cards with the experimental groups. The control groups followed the institutions prescribed vocabulary course which did not include the use of word cards. Additionally, this researcher found no studies seeking the views of Arab learners on the usefulness of word cards. This gap in the literature was addressed through soliciting the students' perceptions during focus group interviews and a survey questionnaire.

The three specific objectives were to:(1) Investigate how decontextualised vocabulary study, using word cards and translation, contributed to a gain in receptive vocabulary for elementary

level Emirati learners of English; (2) Investigate how vocabulary size is correlated with reading comprehension scores among elementary level Emirati learners of English, and (3): Explore the perceptions of elementary level Emirati learners of English regarding the teaching and learning of vocabulary and its relationship to reading comprehension. The philosophical stance of the researcher was vindicated, because the mixed methods research design, underpinned by constructive realism or pragmatism, provided quantitative data that was enriched and corroborated by qualitative data. Despite its limitations, the main conclusions were that (a) decontextualised vocabulary study, using word cards and translation, contributed a more rapid gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element; (b) the size of the receptive vocabulary appeared to correlate with reading comprehension scores. This correlation was especially strong in the case of the Preliminary English Test (PET); and (c) the participants in the experimental group perceived that word cards and translation was a very effective approach to learning vocabulary. The practical implications were that decontextualised vocabulary study, using word cards and translation, could potentially be introduced into curriculum, in order to contribute to a gain in receptive vocabulary for elementary level Emirati learners of English. The findings of this study underline the importance of improving vocabulary size in the case of elementary learners and that the learners are likely to engage better with strategies they believe in.

ACKNOWLEDGEMENTS

My PhD journey began late. I had been a teacher and teacher trainer for 25 years before embarking on it. My frustration with some aspects of teaching and learning and a desire to throw more light onto a particular language learning issue were my motivation. Approaches and ideas in teaching that my practical experience made me doubt were still peddled as answers when I saw them as the problem. This irritation also proved motivating.

The process has been, for various reasons, more difficult than I originally foresaw. I lost family and friends as I came towards the end of my study. My brother and friend Martin left us far too early in August 2016, my mother Sheila passed away in July 2017 my original supervisor, Stephen Bax, sadly died in November 2017. Perhaps counterintuitively, these setbacks did not decrease my determination. They did, however, interrupt my flow. Stephen, who was with me for seven years, inspired me to look at the bigger picture and his advice and friendship will stay with me.

The completion of this dissertation would not have been possible without the support of Fumiyo Nakatsuhara and Chihiro Inoue. They were my supervisors after Stephen. Their continual prodding and advice kept me going although it probably drained them. Thank you also to Daniel Lam for commenting on my thesis, and to Cara Senouni for helping with administrative issues. Finally, my deep gratitude goes to my wife Caitlin for her unwavering support,

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

Background

Vocabulary knowledge includes not only knowing a word's form (e.g., pronunciation, spelling, and morphology), but it also implies an understanding of how that word fits into the world of language. The review of Stahl and Kapinus (2001) focussing on research on vocabulary knowledge, concluded that that to know a word not only means recognizing that word's definition and its logical relationship with other words, but it also means understanding how that word functions in practice in the real world. Developing an extensive lexicon (i.e., a good vocabulary) is believed to be a key factor in the acquisition of a second language (L2) that is different to the learner's native language (L1). Receptive vocabulary acquisition is believed to have a central role in L2 learning (Butler, Urrutia, Buenger, Gonzalez, Hunt, & Eisenhart, 2010; Duppenthaler, 2007; Folse, 2004; Hunt & Beglar, 2005; Richards, 2000; Zimmerman, 1997). Receptive vocabulary refers to all of the words that an individual knows. In contrast, expressive vocabulary refers to words that a person can produce by speaking or writing (Owens, 2001).

Many different factors must be considered by researchers investigating how receptive vocabulary is acquired, including the extreme variations in the learner's baseline levels of word knowledge; how, why, and which words are taught, the different strategies used to teach vocabulary, and the different tests that have been developed to measure word knowledge (Stahl, 2005). Due to the complex interaction between many factors, the most effective strategies for teaching and learning an L2 vocabulary have been disputed. This controversy is manifested by

three debates in the literature, central to vocabulary teaching and learning, as outlined in the following three paragraphs.

The first debate in the literature focuses on whether incidental or intentional learning provides the most efficient means of building vocabulary size. Hulstijn (2001) suggests that incidental vocabulary learning takes place through activities where the aim is not focused on vocabulary development (e.g., reading), whereas intentional vocabulary learning employs activities and strategies with the objective of memorizing vocabulary. Some researchers argue that learning a large vocabulary must be connected to incidental acquisition through reading (Goodman 1967; Krashen, 1985; 1989, 2002; 2004 2013; Pitts, White, & Krashen, 1989; Schmitt & Carter 2000; Smith 1975; Sternberg 1987). However, the claim that sufficient vocabulary for independent reading can be acquired incidentally through reading has been criticised (Grabe, 2004, Nation & Chung, 2009; Stahl, 2005). The amount of reading required is so large to render any lexicon gains to be painfully slow (Cobb, 2008). Empirical evidence indicates that an explicit focus on vocabulary learning is required if vocabulary size is to grow at a rate which would allow elementary L2 learners to quickly reach a threshold adequate for independent reading (Hunt & Beglar, 2005).

The second debate in the literature central to vocabulary teaching and learning is whether words are best presented in context (contextualized) or out of context (decontextualized). Some researchers (e.g., Krashen, 1989; Nagy, 1997) have proposed that the study of decontextualized lexis is unnecessary or inadequate for helping L2 learners acquire the large vocabulary that they will ultimately need (Nagy, 1997). Many advocates of contextualized learning (Twaddell, 1973; Corrigan, 2007; Gardner, 2007; Webb, 2008) argue that contextualized learning is a necessary

condition if more than a superficial understanding of the target language is to be attained, yet there remains fierce debate, with other researchers (Nation, 1985; Prince, 1996; Laufer, 2003; Nation, 2003) suggesting that, whilst contextualized learning is useful at higher L2 levels and in L1 contexts, lower level L2 learners are better served by decontextualised methods.

The third debate in literature concerns the use of translation from the L1 language to English as a strategy for learning vocabulary. The use of the learner's L1 remains taboo to many teachers of English as a foreign language (Auerbach, 1993; Cook, 2010; Hall and Cook, 2012). Phillipson, 1992; Ramachandran, Devi, & Rahim, 2004). In contrast to the teaching profession, some learners of English as a foreign language appear to have faith in the value of translation from L1 to L2, including Taiwanese students (Liao, 2006). A study of low-level English foundation students at a United Arab Emirates (UAE) university sector institution also revealed that the students were generally in favour of translating Arabic to support English language learning (Mouhanna, 2009).

The current research, focusing on vocabulary learning and reading comprehension among a sample of elementary level Emirati learners of English was conducted in the light of the controversial historical background outlined above. The numerous debates in the literature contributed directly to the formulation of the following problem statement, aims, and objectives.

Problem Statement

Vocabulary is a difficult problem for elementary English as a foreign language learners (EFLs). The extent of EFLs knowledge of English vocabulary is only a fraction of the knowledge of native speakers of English. The failure to understand even a few words of an English text can have negative effects on the reading comprehension of EFLs (Gandara,

Maxwell-Jolly, & Driscoll, 2005). Effective L2 vocabulary acquisition is particularly important for EFL learners who frequently acquire impoverished lexicons despite years of formal study (Hunt & Beglar, 2005). However, there is limited previous research focusing on the vocabulary size and reading comprehension of elementary level Emirati learners of English. Based on an extensive review of the literature, in the context of the current study, the three main gaps in knowledge appear to be as follows:

1. How decontextualised vocabulary study and translation may contribute to the learning of vocabulary among elementary level Emirati learners of English;

2. How receptive vocabulary size correlates with reading comprehension among elementary level Emirati learners of English;

3. The reasons why elementary Emirati learners of English perceive that what is demanded of them in terms of achievements in vocabulary and reading is very difficult in practice,

In response to (1) above, Ramachandran and Rahim (2004) and Kuo and Ho (2012) supported the use of decontextualized vocabulary study, using word cards and translation, to improve the size of the receptive vocabulary of learners of English as a foreign language. In response to (2) above, the vocabulary required in the International English Language Testing System (IELTS) reading comprehension test is known to be very difficult for Arab learners of English (IELTS Annual Reviews, 2006, 2007; O'Sullivan, 2009; Cambridge ESOL research notes, 2010). In response to (3) above, research in the UAE has highlighted the many issues facing students in the foundation years of English medium higher education institutions (Davidson, Atkinson & Spring, 2011). Elementary level Emirati learners of English have

generally not been successful and may have developed a perception that what is demanded of them in terms of English vocabulary and reading cannot be achieved, including achieving high scores in the Vocabulary Levels test (VLT); the Vocabulary Size Test (VST), the Academic Word List (AWL), the International English Language Testing System (IELTS) test and/or the Cambridge Preliminary English Test (PET). Consequently, more classroom-based studies, which aim to discover the most efficient approaches for building English vocabulary, based on the test scores and perceptions of Emirati students in higher education institutions in UAE have been called for (Watts, 2011). There is a need to investigate the effect of translation, which has long been criticized in in the UAE, as a tool for vocabulary learning (Schmitt, 2008).

Aims and Objectives

The overall aim of this study was to provide new knowledge and understanding regarding vocabulary learning and reading comprehension among elementary level Emirati learners of English. The three specific objectives of this study were to:

1. Investigate how decontextualised vocabulary study, using word cards and translation, contributed to a gain in receptive vocabulary for elementary level Emirati learners of English.

2. Investigate how vocabulary size is correlated with reading comprehension scores among elementary level Emirati learners of English

3. Explore the perceptions of elementary level Emirati learners of English regarding the teaching and learning of vocabulary and its relationship to reading comprehension.

Research Questions/Hypotheses

The three research questions that guided this study were as follows:

RQ1: To what extent, and in what ways, does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element?

RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores?

RQ3: What are the perceptions of elementary level Emirati learners of English regarding the learning of vocabulary and its relationship to reading comprehension?

The three stated research questions fulfilled the essential criterion that they could not be answered simply by "Yes" or "No". A question that is syntactically orientated toward a "Yes" or "No" answer (e.g., Does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element?) has limited logical, pragmatic, or scientific value. A question that can be answered only by "Yes" or "No" does not provide for complex answers that reflect the infinite shades of grey that exist within the extremes of black and white (Hurley, 1992; Mahmood, 2014; Radnitzky & Anderson, 1978; Tindale, 2007).

The two research hypotheses tested in this study, linked to RQ1 and RQ2, were as follows:

H1: Decontextualised vocabulary study, using word cards and translation, contributes toward a more rapid gain and a greater amount of receptive vocabulary among elementary level Emirati learners of English, than a similar teaching programme lacking this element.

H2: The receptive vocabulary size of the elementary Emirati learners correlate with the PET reading scores. In contrast, the IELTS reading scores correlate only with the receptive vocabulary size of those participants who exhibited the greatest receptive vocabulary gains.

As discussed in the instruments section of chapter 3, the IELTS academic reading module would not appear to be an ideal instrument for use in a study with participants at the level of those in the present study. It seemed likely that the PET reading tests would prove to be a more appropriate instrument, because the PET reading texts are controlled for vocabulary at a level that may be accessible for these students, whereas the far greater vocabulary load found in the IELTS reading texts would prove overwhelming. However, IELTS was mandated by the institution as the 'gatekeeper' for entry to the undergraduate programme and, therefore, the participants were required to prepare for and sit the examination.

H1 and H2 are research hypotheses, because they are statements created by a researcher speculating upon the outcome of research (Shuttleworth & Wilson, 2008). H1 and H2 fulfil the definition of a hypothesis as 'a proposed explanation that has not been tested before, made on the basis of limited evidence, and used as a starting point for further investigation' (Bowker & Randerson, 2006, p. 21). H1 was based on limited evidence regarding the efficacy of word cards and translation for elementary level Emirati learners of English. H2 was based on limited evidence regarding the analysis of IELTS and PET test scores. Because H1 and H2 have never been tested before, they were applicable as starting points for further investigation.

Theoretical Framework

The process of answering the research questions and testing the hypotheses was underpinned by the philosophical stance of the researcher, defined as constructivist realism (Cupchick, 2001) otherwise known as pragmatism (Creswell, 2014). This particular philosophical stance meant that the researcher believed that objectivism and social constructivism are not polarized but are compatible. Objectivism posits that knowledge consists of an objective and sacrosanct body of information, separate to the human mind, so that facts and human feelings are not connected (Giddens, 1974). Social constructivism posits that knowledge is a complex variable that is affected by human feelings and is negotiated and rationalized through social interaction. Consequently, "reality is determined by people" (Easterby-Smith, Thorpe, & Lowe, 2002, p. 32) and language is the essential system through which people interact to construct knowledge (Amineh & Asl, 2015; Palinscar, 1998). Social constructivism is supported by the history of science, which has demonstrated that previous knowledge, including widely accepted paradigms, (e.g., the earth is flat) was contradicted and discarded after new information was discussed and accepted by the scientific community, leading to the re-construction of existing knowledge (Kuhn, 2012).

In the context of educational research, objectivists tend to use deductive reasoning to interpret quantitative data, whereas social constructivists tend to use inductive reasoning to interpret qualitative data (Fraenkel & Wallen, 2011). The researcher, however, did not support the polarity of objectivism and social constructivism. He believed that the collection and interpretation of a combination of quantitative and qualitative data was essential in order to construct new knowledge regarding the vocabulary learning and reading comprehension of

elementary level Emirati learners of English. Constructive realism or pragmatism, which involves a combination of objectivism and social constructivism enabled the building of bridges between quantitative and qualitative methodologies, and facilitated the answering of the complex research questions. Cupchick, 2001; Creswell, 2014)

Research Design

The philosophical stance of the researcher was aligned to a mixed methods sequential explanatory design, whereby quantitative data, collected by means of a survey and an experiment, was followed up by the collection of qualitative data, collected by means of interviews (Creswell, 2014). The reason for implementing this particular research design was that complex research questions concerning the impact of educational interventions need to be answered pragmatically by means of the most useful tools available, involving the integration of quantitative and qualitative methodologies (Fraenkel & Wallen, 2011). The use of a mixed methods sequential explanatory design implied that no barriers existed between quantitative and qualitative methodologies, empowering the researcher to explain the meaning of quantitative data in more detail through the interpretation of qualitative data (Bryman, 2007; Creamer, 2018). The results of the statistical analysis of the quantitative experimental and survey data were therefore enriched by the interpretation of qualitative data, including the participants' perceptions of the quality of the teaching and testing methods employed, as well as the difficulties they experienced in learning English vocabulary.

Scope and Delimitations

The scope of this mixed methods study was limited to an investigation of the vocabulary learning and reading comprehension of a sample of elementary level Emirati learners of English during two semesters in the 2010/11 academic year. The study assumed that limited quantitative and qualitative data currently exists regarding the effectiveness of the use of word cards and translation as an educational strategy for elementary level Emirati learners of English. The collection and analysis of data were restricted to the quantitative responses to a cross-sectional survey, the scores achieved using the VST, VLT, AWL, IELTS and PET tests, and the qualitative responses to interview questions.

Significance of the Study

The findings of this study are significant because they may help English language teachers and administrators in the UAE to make educational policy decisions in the future that will benefit the acquisition of vocabulary by elementary Emirati learners of English. The findings of this study may support a policy to implement decontextualized vocabulary study, using word cards and translation as an effective educational intervention in the classroom, specifically for elementary level Emirati learners of English in the UAE. The findings may also justify the introduction of decontextualized vocabulary study in the curriculum for students who are participating in courses to train as teachers of English as a foreign language. Alternatively, the findings of this study may indicate that decontextualized vocabulary study, using word cards and translation is not an effective educational intervention, and English language teachers should therefore be advised not to apply this strategy in the classroom.

Organisation of Thesis

The remainder of this thesis is organized as follows: Chapter 2: Literature Review; Chapter 3: Methodology; Chapter 4: Results. Chapter 5: Discussion and Conclusions.

CHAPTER 2: Literature Review

Introduction

This chapter reviews the key literature on the treatment of vocabulary learning in a second language and the relationship between vocabulary knowledge and reading. The present study focuses on these areas in the specific context of elementary level Emirati learners of English. Meara (2002) welcomed the renewed interest in vocabulary that had become evident in the previous decade but argues that researchers from the field of linguistics needed to broaden the scope of their work by considering the theories emerging from other fields; most significantly computational linguistics and psycholinguistics. However, although his concerns over the dangers of a disjunction between these related theories seem perfectly valid when we review the missed opportunities for vocabulary teaching in the past, today, with the benefit of hindsight, the situation appears less worrying. Nation (2003) highlighted why Meara might justifiably be concerned. He claimed that research arguing against the use of authentic materials; top-down approaches to teaching reading; and incidental vocabulary learning had been available for fifty years but had been ignored by English Language Teaching (ELT) publishers and materials writers With the above in mind, the first section of this chapter investigates how vocabulary has been situated historically in the various approaches and methods that have been employed in second language teaching in order to provide a pedagogical context for why this study is needed.

Next, as this study is primarily concerned with learning vocabulary, the focus moves to defining what knowing a word means. Different types of word knowledge are described and the types most important to this study are identified. This is followed by a section on the construct of reading ability. The correlation between vocabulary and reading for elementary Emirati readers

of English is another key aspect in this study. The components involved are listed and discussed and the components most important to the present study are identified.

The following section focuses on the key debate of whether vocabulary should be taught intentionally or acquired implicitly. Firstly, the origin of these terms is considered and they are then defined in the context of this research. The section then goes on to highlight a further important question in vocabulary learning; whether words are best learnt in or out of context. Next research studies on decontextualised methods of teaching vocabulary are examined including a specific focus on the use of translation and word cards. This section considers studies from the field of linguistics, particularly Second Language Acquisition (SLA) and Psycholinguistics and underlines the complementary evidence on the benefits of retrieval these disciplines provide. This section concludes by investigating how much recycling of vocabulary is required to ensure words will be retained; recycling is considered from the viewpoints of both vocabulary size and depth (Nation, 2008). Materials are evaluated in relation to which vocabulary is likely to be met and how often words are encountered

The focus next moves to incidental vocabulary learning in the modern era. This section reviews the roots of incidental learning and considers studies investigating vocabulary acquisition through reading and, in one case, through listening. This is followed by articulating and examining a number of assumptions made by proponents of vocabulary acquisition through reading. The following section discusses the performance of Arab learners when reading in English. It then describes the particular problems Arab students encounter in recognizing English words. The next section focuses on studies aimed at identifying a vocabulary size threshold which would provide adequate text coverage for independent reading. These are examined in

some detail with the aim of making the task faced by the participants in this study clear. Finally, beliefs about vocabulary learning strategies are reviewed from the perspectives of theorists, practitioners and language learners.

Vocabulary: Overview

After decades in the wilderness vocabulary is now once again seen as having a central role in language learning, at least as far as many researchers are concerned (Quian, 1993, Duppenthaler, 2007). This has often not been the case in the past. Richards (1976,) claimed that:

> The teaching and learning of vocabulary has never aroused the same degree of interest within language teaching as have such issues as grammatical competence, contrastive analysis, reading, or writing, which have received considerable attention from scholars and teachers. The apparent neglect of vocabulary reflects the effects of trends in linguistic theory, since within linguistics the word has only recently become a candidate for serious theorizing and model building. (p. 77)

This does not mean that there has been a complete lack of focus on vocabulary rather that vocabulary has never been the driving force of a dominant methodology. In this section the treatment of vocabulary through the history of language teaching and learning is reviewed with the aim of providing an understanding of the past and a context for vocabulary treatment in the present study. We begin with the Grammar Translation Method (GTM) which was developed as an attempt to adapt the earlier Classical Method to teach modern languages to large classes (Zimmerman, 1997). The Classical Method was employed to learn Latin and Greek, through the

reading of classical literature in the original script, and counted the provision of mental exercise to develop the intellect amongst its aims. Clearly, there was no focus on spoken communication as these languages were no longer spoken. The dawning of the industrial revolution and the increased possibilities for travel and commerce brought a realization that it might be useful to teach living languages with the goal of better communication with foreigners. Whereas previously language study had mainly been the province of scholars engaged in solitary study there was now a need to teach large classes in a school setting. Howatt (1984) made the point that the GTM was originally conceived as a solution to this issue:

> The grammar-translation method was an attempt to adapt these traditions [of the Classical Method] to the circumstances and requirements of schools. It preserved the basic framework of grammar and translation because these were already familiar both to teachers and pupils from their classical studies. Its principal aim, ironically enough in view of what was to happen latter, was to make language learning easier. The central feature was the replacement of traditional texts by exemplificatory sentences. (p.131)

Richards and Rogers (1986) state that the GTM became the favoured language teaching approach in the 1840's and maintained this position until the 1940s. Indeed, in many parts of the world it is still dominant (Bamford 1993, Chang 2011, Mondal 2012). That the GTM became popular, and has managed to remain so in some contexts, may be due in part to the relative ease of lesson preparation and that it requires a modest investment in teacher education.

The key characteristics of the GTM include deductive learning of grammar rules; applying these rules when translating texts; most teaching is conducted through L1; vocabulary teaching through direct translation; bilingual dictionaries and bilingual word lists. However, as

the name makes clear, the GTM primarily focused on grammar with a subservient role for vocabulary. This had serious implications as to how and which vocabulary was taught. Vocabulary was chosen when it was useful in illustrating a grammatical rule, with no weight given to how frequent the words might be or if they might be useful in spoken communication (Kelly 1969). It should also be noted that much of the reading in the GTM consisted of literary texts which were unlikely to yield vocabulary for everyday use (Rivers, 1981).

Although, as noted above, the GTM was dominant in language teaching for a considerable period and still survives to this day it has always attracted criticism. One early critic was Prendergast (1864/2013) who seized upon the lack of focus on useful vocabulary in the method. He called for the creation of new vocabulary lists, based on word frequency, to replace the lists of often archaic language then in use. This view failed to attract sufficient support to make an impact at the time but is a remarkable forecast of what is once again seen as an important aspect of vocabulary acquisition today.

A further, more significant, challenge came with the Reform Movement in the late 19th century. From the reformers' viewpoint the focus of language learning should primarily be on the spoken word and learning was to begin with study of phonetics followed by a grammatical stage. Although some vocabulary was seen as necessary it was relegated to a minor role with no study of individual words. Sweet (1899/1964), a leading reformer, makes this clear from a phonological stance, He contends that basis of language is the sentence and that, in phonetics, the word alone has no place.

Kelly (1969) points out that Sweet did encourage the learning of vocabulary which might prove to be the most practical. He also foresaw that wordlists, based on frequency and produced using statistical procedures, could be important tools in language learning, although this insight was not followed up until much later and by other theorists.

The Reform Movement's attack on the GTM paved the way for the Direct Method, one of the alternative approaches proposed at the time. There does not seem to be any one individual behind the Direct Method, although its roots can be traced to teaching in German State schools in the 1880's. Richards and Rogers (1986) note that the experts of the day did not consider Direct Methodology to have a fully developed theory despite the fact its adherents insisted on some basic principles; the main focus should be on the speaking skill; the language of instruction should be the target L2; and translation should not be used. Vocabulary is taught by using realia and demonstrations, albeit only for words with concrete meanings, whilst abstract vocabulary was meant to be linked to ideas (Zimmerman, 1997). Brown (1973) found this treatment of vocabulary to be a major weakness in the method because of the time spent conveying vocabulary meaning and the great difficulties in doing so. The Direct Method was also known as a 'Natural Method' as it attempted to apply what was thought to be how L1 was acquired to how L2 could be taught, unfortunately without a full understanding of L1 acquisition and little consideration of the differences between that and learning a second language (Schmitt, 2002). Interestingly, a pattern now begins to emerge; proponents of the new Direct Method rejected all the GTM had to offer. This was to be repeated at other crucial stages in L2 language teaching methodology development and often without sound reasons. Carter and McCarthy (1988) noted that:

> The history and development of vocabulary teaching is not so much one of old insights leading to new; it is more a series of dominating ideologies or fashions that have succeeded one another and which sometimes come full circle. (p.39)

This may be especially significant when we consider the next important developments and the paths taken subsequent to these.

West, remembered today for his General Service List (1953) was a pioneer not only for his work with word lists. He was also a leader in the Reading Approach Method and created a framework for learning to read which had vocabulary at its heart. West was a colonial educator, who spent around twenty years in India where he was involved in teacher education and writing course books. He was unusual in that his research was based on what he saw in the classroom and could be described as what is now known as 'action research'. West (1926) advocated a bilingual approach to the use of English in India, where the main goal would be to teach the students how to read in order to open up the wealth of reading materials in English. He argued that speaking English was not necessary or attainable for the majority of students and that even where complete mastery of the language was the aim, reading should lay the foundation. West (1926-27) produced 'The New Method Readers', a reading course underpinned by his research and seminal thinking. In today's climate of renewed interest in vocabulary and reading, even a cursory study of West's work at once makes clear the originality of his thought and the opportunities that may have been lost in not building on his ideas. West's reading course encompassed many ideas now once again in vogue and exemplifies Carter and McCarthy's quote above. In summary, the course focused on the most useful vocabulary taking into account

frequency and range; it included graded readers as supplementary materials in order to recycle vocabulary; new word density in the revised series was 1 in 60 running words equating to 98.3% known words (West, 1927); each level ended with a further reader containing no unknown words; early lessons included very short stories written with as few as 100 words allowing very low-level learners to begin reading; top-down skills were not ignored but were left until a solid foundation had been built; the course was designed to rely on the books rather than the teacher, in order to lessen the impact of variable teaching skill on learning (West, 1960).

The Reading Method remained popular in the United States until the early 1940s and for considerably longer in India but a combination of factors led to a renewed focus on Oral-Aural skills. The Audio-lingual Approach, which dominated from the late 1940's until the Seventies, had its origins in the USA's need for speakers of foreign languages during World War 11. Richards and Rodgers (1986) describe the successful results of the Army Specialized Training Program, although they argue that this had little to do with any breakthrough in methodology rather that it relied on the intensive training of small groups of mature, dedicated students. However, the relative success of the two-year programme generated interest in the linguistic community and led to Audio-lingualism. The first classes using the approach took place at the University of Michigan, where the structural linguist Charles Fries was director of the language institute. The approach saw language structures and an oral-aural focus as the fundamentals of language learning. Structures were identified and drilled repetitively until they were retained; language learning was seen as habit formation (Decarrico, 2001). This structural theory was complemented and underpinned by the behaviourist psychology of Skinner (1957) which saw all behaviour, including communication, as being learnt through habit formation. Fries (1945) made

it abundantly clear that vocabulary was to play a subsidiary role, cautioning that students should not equate word knowledge with language learning. Vocabulary was introduced only when necessary for facilitating the practice of a structure.

The Audio-lingual Approach began to attract heavy criticism in the 1970's leading to a further paradigm shift. This shift can be illustrated by the theoretical changes found in the work of leading figures of the time; here the work of Rivers is cited, although this is not meant to convey she was alone in changing tack. Rivers (1968), echoing Fries (1945) suggested that too much vocabulary was not good for low-level language learners as they may come to believe that this should be the most important element in their studies. In contrast, Rivers argued that audio-lingualism had produced learners who could perform classroom drills competently but were bereft of the skills needed to communicate outside of the classroom. She now considers that practitioners should focus more on word meanings in order to promote fluency. Earlier, Wilkins (1972) established a new model of language systems which can be viewed as supporting the importance of vocabulary. He proposed two categories of meaning; functional (invitations, requests) and notional (ideas such as 'time'). Laufer (1986) suggests that syllabi based on these concepts are likely to focus more on lexis than grammar given the importance of themes and situations in such syllabi. Further support came from Twadell (1972) who called for the development of vocabulary size at intermediate level and suggested that this be achieved through learners employing guessing strategies given that insufficient classroom time was available. Wilkins (1974) solution to learning sufficient vocabulary involved making sure the learner was exposed to huge amounts of the target language. With hindsight, although this period sees an acceptance of the importance of vocabulary, the suggestions made on how to focus on it seem

vague and impractical. Indeed, given the renewed acknowledgement of vocabulary's importance, it is perhaps surprising that little theoretical focus on how vocabulary should be taught is found in the literature on notional/functional syllabi.

The work of Wilkins and others in redefining language as system of meanings behind communication, rather than categories of grammar and vocabulary, was one of the foundations for Communicative Language Teaching (CLT). Hymes (1972) can be seen as making this redefinition more concrete when positing 'communicative competence' as an alternative to Chomsky's (1965) seminal theory of linguistic competence. Hymes built on and added to Chomsky's foundations of competence and performance; he drew a distinction between Chomsky's contention that linguistic theory should focus on an idealised speaker/listener in a homogeneous speech community and his own theory which includes sociocultural perspectives. Hymes asserted that linguistic competence is one of many factors included in communicative competence and that other factors, such as; social class; individual ability; opportunity to learn; and motivation affect the degree to which communicative competence is attained. Communicative competence in this view includes both linguistic knowledge of a language and the ability to use a language; without knowledge of social and cultural values informed use is impossible. In a similar vein Widdowson (1978) labels the ability to display linguistic knowledge as 'usage' and the ability to communicate effectively as 'use' with both being aspects of performance; he illustrates how it is perfectly possible to possess linguistic knowledge but not be a skilled communicator. As we can see, the advent of CLT firmly places the focus on language as communication rather than one of learning abstract linguistic rules, form is now to be learnt through a primary focus on expressing meaning.

It should be noted that CLT is seen as an approach rather than a method. Anthony (1963) proposed that an approach consists of a particular theory of the nature of language and language learning whereas a method can be seen as the plan used to implement an approach. As Anthony suggests, any one approach can spawn a number of methods and the CLT approach is not an exception. What these methods should have in common is a clear correlation with the principles of the approach. In the case of CLT the principles of the approach include the primacy of fluency over accuracy; the belief that real communication is more likely to produce learning than the repetitive, inauthentic practice found in Audio-lingualism; the principle that language should be used to work on authentic tasks and that the learning activities should be designed to engage the learner (Littlewood, 1981; Stern, 1981; Rivers, 1983).

Given that the ability to communicate effectively is highly valued in CLT, it is tempting to assume that vocabulary would be seen as central in the approach. Indeed, Wilkins (1976) takes the view that whilst a lack of grammar is an obstacle to communication a lack of vocabulary makes it impossible. Widdowson (1987) asserts that understanding is possible with an imperfect command of grammar but not without knowledge of vocabulary. Krashen (1982) suggests a greater focus on vocabulary in order to help learners understand messages. Surprisingly, any such assumption would be erroneous. Schmitt (2002) notes, that in common with previous approaches, no direct attention is given to vocabulary. Krashen (1982), in keeping with his views on the superiority of acquisition over learning, argues that if there is enough access to comprehensible input vocabulary might well take care of itself. Once again vocabulary is assigned a secondary role, this time in support of functional language use (Decarrico, 2001).

The corpus-based work on a lexical syllabus led by Sinclair seemed to herald a breakthrough in vocabulary teaching (Sinclair, 1987; Sinclair & Renouf, 1988). The newly available research possibilities, allowed by rapid computation of language, not only confirmed previous estimates of word frequency but uncovered compelling evidence of how words occur in habitual patterns. This work was part of the joint COBUILD project between Collins publishers and the University of Birmingham. The research informed the Collins COBUILD English Course (CCEC), (Willis & Willis, 1988) although this followed a hybrid syllabus rather than a purely lexical one (Long & Crookes, 1993a)**.** Vocabulary load is controlled; Level one 700 words; Level two 1500 words; Level three 2500 words but is not clear how much it is recycled and there is also some focus on grammar. Willis (1990) claims that the vocabulary from each level is recycled in the next without saying how often. However, he admits that the syllabus does not dictate what will be learnt and proposes that, when working on tasks, students should focus on achievement rather than any specific vocabulary or grammar. It is questionable how systematic vocabulary learning will take place in what is essentially a task-based approach with traces of a structural syllabus added. Long and Crookes (1993a) point out that the map for level one includes titles suggesting a little of all approaches rather than a genuinely lexical one, and that where there is a focus on lexis it is generally post task thereby sacrificing context. Parana (1998) claims that some of the target vocabulary in the CCEC level one is not found in the texts and, more surprisingly, only 500 words are truly the most frequent whilst others were chosen for their value when writing course materials. Notwithstanding these criticisms, this course book series arguably assigned more weight to vocabulary than any since West's (1926-27) New

Method Readers. The CCEC is now out of print due to low sales and a failure to gain a large following amongst teachers.

The Lexical Syllabus was followed by the Lexical Approach which appeared in the early nineties. Lewis (1993) outlined 'the way forward', which was the sub-title of The Lexical Approach, in his challenge to existing methodology. Yet on closer inspection his new approach may not be as radical as Lewis would have it appear. He proposed a syllabus based around the idea of language as lexical chunks as a replacement for the dominant grammar driven syllabi of the time. The PPP (presentation-practice-production) teaching method was to be superseded by OHE (observe-hypothesise-experiment) with noticing activities as the mechanism for prompting acquisition (Thornbury, 1998). Input here is all of the language that a learner is exposed to whilst intake is the part of this language understood by the learner. Thus, in order to facilitate acquisition, teachers following a lexical approach need to ensure that learners notice language that is important. Unfortunately, Lewis does not make clear how this is to be achieved; he seems to offer techniques without a consistent theory. On the one hand, he appears to be in agreement with Krashen (1985) on the necessity of learners meeting vast amounts of comprehensible input (Lewis, 1993) but on the other, at odds with him because of the emphasis on noticing in his lexical approach (Lewis, 1997), an explicit method of learning not in line with Krashen's position. Lewis (1993) famously declares his view on language: 'Language consists of grammaticalised lexis not lexicalised grammar' (p.vi) but is not so clear on which of this grammaticalised lexis should be taught. Hill and Lewis (2002) produced a dictionary of collocations but did not suggest which might be more valuable to the learner; Lewis (1993) seems to suggest that with sufficient recycling and exposure to input, learners will somehow

acquire these chunks but given the amount and size of the chunks this seems doubtful. It should also be noted that Schmidt's (1990; 1993a; 1993b; 1995) noticing hypotheses has attracted criticism. Truscott (1998) questions whether there is any empirical evidence to support Schmidt's claims and suggests that due to the vague nature of the hypotheses any effect would, in any case, be difficult to measure. In light of the contradictions in the Lexical approach as proposed by Lewis, it might best be viewed as useful in questioning existing paradigms and focusing the spotlight on vocabulary without providing coherent answers to the issues raised.

In the present era a key vocabulary research area focuses on identifying the amount of words a learner needs to know in order to tackle academic texts. Early attempts to answer this question suggested a vocabulary size of 5,000 word families giving 95% text coverage might suffice (Laufer, 1989). However, there is now a consensus that these figures are too low and that 8,000 to 9,000-word families giving 98% coverage are needed to read these texts efficiently and at a reasonable pace (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt, Jiang & Grabe, 2011). This question is explored in greater depth below.

A second area for contemporary studies discussed below stems from the question above. If we accept that an 8,000 to 9,000 vocabulary size is required for academic study it raises the question of how second language learners, who wish to study in an English medium institution, can achieve this size in an economical manner given the constraints on their time (Schmitt & Schmitt, 2014).

A third important development in vocabulary acquisition study during recent years has been the illumination provided by other disciplines; the fields of Psychology and

Psycholinguistics appear to be particularly important in this respect and their influence is discussed below.

In summary, this section has demonstrated the changes in importance assigned to vocabulary in language teaching throughout some of the key stages in its history. At various stages researchers have stressed the key role of vocabulary yet it has seldom occupied a central position in the classroom. In the past twenty-five years there has been a massive resurgence in researcher interest. However, whilst there is a large and growing body of evidence which suggests that vocabulary should play a far more prominent role in second language teaching and learning, the pedagogical implications of this research have still not been implemented in many language classrooms or reflected in learning materials (Folse, 2010; Matsuoka & Hirsch, 2010; Nation, 2003; Schmitt & Schmitt, 2014).

Types of Vocabulary Knowledge

If it is true vocabulary learning should play a greater role in language learning it would be wise to identify what learning words means. This issue has been approached from the learners' viewpoint using rating scales from 'I have never seen this word' (no knowledge) to 'I can use this word in a sentence' (operational knowledge), (Paribakht & Wesche, 1993). Whilst this is a useful starting point it does not make clear what aspects of the word have been learnt on the journey from no knowledge of the word to use of the word. Other researchers have taken the approach of identifying distinct elements of word knowledge in order make clear what fully knowing a word might involve (Qian, 1999; Read, 2004; Richards, 1976). Schmitt (2014) claims that table 1 is still currently thought of as the most complete attempt at specifying all aspects of word knowledge.

Table 1

Guidelines for Word Knowledge (R=Receptive; P=Productive)

Meaning	Form and meaning	R What meaning does this word form signal? P What word form can be used to express this meaning?
	Concepts and referents	R What is included in the concept? P What items can the concept refer to?
	Associations	R What other words does this make us think of? P What other words could we use instead of this one?
Form	Spoken form	R What does the word sound like? P How is the word pronounced?
	Written form	R What does the word look like? P How is the word written and spelled?
	Word parts	R What parts are recognizable in this word? P What word parts are needed to express this meaning?
Use	Grammatical functions	R In what patterns does the word occur? P In what patterns must we use this word?
	Collocation	R What words or types of words occur with this one? P What words or types of words must we use with this one?
	Constraints on use	R Where, when and how often would we expect to meet this word? P Where, when and how often can we use this word?

Source: Nation (2001, p.27).

A language learner contemplating the challenge detailed above might understandably consider that learning a word was an almost impossible task. Fortunately, for our imaginary learner, learning a word is not an all or nothing task. Nation's specifications identify both receptive (R) and productive (P) elements of word knowledge. Elementary level learners of English are likely to learn the receptive elements initially, and it is some of these elements that

lend themselves to the type of intentional learning investigated in this study (Schmitt, 2014). The form-meaning link, which is tested through the vocabulary tests employed in this study (Kremmel & Schmitt, 2016), is generally the first stage on the path to knowing a word and can be explicitly taught as in the present study.

Aspects of word knowledge are learnt incrementally, learners may be able to identify the meaning of a written word but this does not mean they could identify the spoken form. Complete mastery of all aspects of word knowledge is complicated and even native speakers learn new meanings and uses of vocabulary they already know in other ways (Schmitt, 2007). The present study is concerned with the automatic recall of the receptive form-meaning link in the context of elementary level learners.

The Construct of Reading Ability

The present study is concerned with any correlation between increased receptive vocabulary size in elementary Emirati learners of English and improvement in reading ability. In this respect it is important to clarify what is meant by 'reading ability'. As with word knowledge in the previous section, reading ability is a complex issue. The reading ability construct involves a number of component parts; can the reader recognise the word; can the reader read quickly etc. Grabe and Jiang (2014) synthesise a number of research articles on the components of reading ability. Their conclusions are outlined in Table 2.

Table 2

The Components of Reading Ability

1. Efficient word recognition processes (automaticity)
2. A large recognition vocabulary (vocabulary size);
3. Efficient grammatical parsing skills;
4. The ability to formulate the main ideas of a text;
5. The ability to engage in a range of strategic processes while reading more challenging texts (including goal setting, academic inferencing, monitoring);
6. The ability to recognize discourse structuring and genre patterns, and use this knowledge to support comprehension;
7. The ability to use background knowledge appropriately;
8. The ability to interpret text meaning critically in line with reading purposes;
9. The efficient use of working memory abilities;
10. The efficient use of reading fluency skills (including reading speed);
11. Extensive amounts of exposure to l2 print (massive experience with l2 reading);
12. The ability to engage in reading, to expend effort, to persist in reading without distraction, and achieve some level of success with reading (reading motivation).

(Modified from Grabe & Jiang, 2014, p.188)

Mastery of all of the components listed here would be in the province of an advanced reader but clearly not of elementary learners. However, as with word knowledge, these skills are not gained at the same time or at a single proficiency level (Grabe & Jiang, 2014). At elementary level the first two components are of critical importance because, without efficient word recognition, the reader will use most of their working memory capacity on attempting to decode partially known or unknown words leaving little for more complicated undertakings (Birch, 2007). In other words, it may be argued that developing automaticity of word recognition and building vocabulary size are important goals for less skilled readers (Khalifa & Weir, 2009). The focus of the present study, involving elementary Emirati learners of English, is therefore on these two components of the reading construct. The following section investigates theories on how vocabulary size may be best built.

Intentional versus Incidental Learning

Whilst there are many aspects to learning a language, research has made it clear that vocabulary is a key factor in second language acquisition, (Folse, 2004; Richards, 2000; Zimmerman, 1997). However, what is disputed is how best vocabulary should be taught and learnt. The main debate centres on whether incidental or intentional learning provides the most efficient means of building vocabulary size; therefore, it is helpful if these terms are clearly defined. Hulstijn (2013) notes that these terms first appeared in the 1950's in the field of behaviourist psychology; intentional signified that a human subject knew they were to be tested after a treatment and incidental that they did not. The meaning attached to the two terms in present day SLA terminology is less clear cut and the sometimes synonymous or at least overlapping use of implicit acquisition and explicit learning further muddies the waters. Again the terms implicit and explicit were originally borrowed from psychology where the definition assigned was that implicit learning took place without the learner being conscious of it, whilst explicit learning is characterised by consciousness (Rieder, 2003). This clearly would present great difficulties for studies where the aim is to measure the difference in vocabulary acquisition when two or more teaching approaches are used. How would the researcher measure what had been learnt or acquired without consciousness? Hulstijn (2001) notes that, in L2 teaching theory, incidental vocabulary learning is said to take place through activities where the aim is not focused on vocabulary development, whereas intentional vocabulary learning employs activities and strategies with the objective of memorizing vocabulary. In this study, these meanings of incidental and intentional learning are assumed. However, it should be noted that this does not imply that incidental learning necessarily occurs in isolation. Ellis (2005) suggests that there is

an interface position between solely implicit or explicit learning. He draws on ideas from the fields of psycholinguistic and neurolinguistics to explain how a conscious explicit focus on language can select the language to be learnt; this is later gradually transferred to long term memory as it is repeatedly implicitly encountered and processed during input. Ellis (2008) argues that this process is the foundation of Schmidt's (1990) concept of 'noticing' in language learning. The scope of my work does not allow for an in-depth analysis of this area though I take account of these ideas in my study.

Contextualized and Decontextualized Vocabulary Learning

A further issue central to vocabulary teaching and learning is whether words are best presented in (contextualized) or out (decontextualized) of context. Many advocates of contextualized learning (Corrigan, 2007; Gardner, 2007; Webb, 2008; Twaddell, 1973) argue that it is a necessary condition if more than a superficial understanding of the target language is to be attained, yet there remains fierce debate, with other researchers (Laufer, 2003; Nation, 1985; Nation, 2003; Prince, 1996) suggesting that, whilst contextualized learning is useful at higher L2 levels and in L1 contexts, lower level L2 learners are better served by decontextualized methods. This appears to be due to the necessity of reaching a threshold level of vocabulary size before useful vocabulary gains can be made through encounters with new words in context. It should be noted that whilst decontextualized work with vocabulary is by nature intentional learning, contextualized work with vocabulary in texts could fall in to both the incidental and intentional categories depending on the primary purpose of the work undertaken.

Decontextualised Learning: Word Cards and the Retrieval Effect

Nation has been an enthusiastic advocate of the use of word cards in vocabulary teaching for many years. Nation (1978) suggests that cards, with the target word on one side and the translation on the other, are a valuable tool in building vocabulary size. Why should this be so? We are given a clue in the seminal work of James (1890) who contended that once something is partly learnt, it is more likely to be fully learnt if we retrieve it from the mind rather than returning to the source material, more recently Roediger and Guynn (1996) defined retrieval as 'accessing stored information' (p.197) from memory. A comparison between the use of word cards and word lists for intentional vocabulary learning exemplifies the point. Word cards, which have the L2 target item on one side and the L1 translation on the other, demand a conscious effort because the language learner cannot see one side of the card. Thus, whether the learner attempts to retrieve meaning, by looking at the target word first and retrieving the translation, or by looking at the translation first and retrieving the form, retrieval from memory occurs if the attempt is successful. In contrast, list learning, where the target item and translation are side by side, requires no comparable mental effort. Nation (2001) argues that word cards with translations appear to be the most efficient method of accelerating the vocabulary growth of low-level learners. He notes that, though there is opposition to this view, the criticisms are not supported by research evidence.

Empirical evidence in support of the benefits of retrieval is provided by a variety of sources. Ramachandran and Rahim (2004) conducted a study involving sixty 16 years old high school students with Malay as their L1; the elementary level participants were divided into experimental (n=30) and control (n=30) groups. The target language consisted of 20 words from

the prescribed course book; these items had been included with other words in a pre-test to ensure that they were unknown to the participants. 5 words were taught each week over a period of 4 weeks with a test of the target words at the end of the week; target vocabulary from previous weeks was tested in weeks 2, 3 and 4. The experimental group were given the translation of each item and these were explained in Malay and English whereas the control group were not given translations and the explanations were in English only. The results showed that the translation group (experimental) recalled the target vocabulary significantly better than the control group on the initial weekly tests. The difference was even more pronounced on a delayed post-test a month later when all 20 items were tested; the authors report a 90% success rate for the experimental group, including a 100% rate for 12 of the items, compared to a success rate of 10% for the control group. Although this study did not use word cards it appears to suggest that, once the relationship between form and meaning was clarified through translation, the frequent recall tests aided retention in much the same manner as testing with cards. The limitations of this study appear to lie in its short duration and the small amount of vocabulary targeted. Elementary level students learning at the rate of 20 words a month or roughly 250 words a year would require many years of study to build a vocabulary size sufficient for independent reading.

Kuo and Ho's (2012) study provides evidence in support of the greater efficiency of word cards over word lists. The participants were four existing groups (n = 120) of Taiwanese 9th grade students with an average age of 15 years. The groups were randomly assigned to be taught under four conditions: (1) word cards with spaced practice; (2) cards with all practice/study done at the same time; (3) lists with spaced practice; (4) lists with all practice/study done at the same time. The participants' previous vocabulary learning experiences had been almost identical with

lists, though not word cards, forming part of the instruction. Pre-testing scores did not reveal any significant differences in vocabulary size between groups. The post-test results revealed significantly superior vocabulary gains for the groups using word cards over those using lists. Although the results regarding spaced versus massed practice did not produce significant differences they did indicate that spaced practice may be more efficient with cards, as the difference between card and list spaced groups was larger than that between card and list mass practice groups. The authors suggest four possible explanations for the higher vocabulary gains of the word card groups: that, as described above, the forced retrieval involved when working with cards leads to better retention; that using cards eliminates the possibility of the list effect where, because the words are always in the same order, one word in the list acts as a prompt to recall of the next; that lists present and encourage a focus on all words in a group even when some have been learnt whereas cards allow students to focus solely on unlearnt words by removing those already learnt from a pack; and finally, that cards may make the learning activity more enjoyable by introducing manual activity and a game type element to the learning process. The beneficial effects of retrieval and/or spacing on vocabulary retention have also been noted by a number of other researchers including Slamecka and Graf (1978), Barcroft (2007) and Karpicke and Bauernschmidt (2011). Kuo and Ho concede that limitation of their study lies in its short eight-week duration. They suggest that a similar study over a longer period might produce different results on retention and, more crucially, allow spaced practice to have more effect.

Royer's (1973) earlier study also investigated the effect of word card testing on vocabulary acquisition. This study involved 3 groups: group 1 used word cards with the English and Turkish equivalents on one side of the card and the Turkish word only on the other, groups 2

and 3 had cards with the English and Turkish words on one side of the cards with the other side blank. Group 1 was asked to study the 20 target words until the individual students felt the items had been fully learnt and also asked to use a self-testing technique. The testing technique followed a period of working with the English/Turkish equivalents in order to get an initial understanding of meaning; the students then engaged in self-testing by looking at the Turkish word only and retrieving the English equivalent from memory. The packs were shuffled after each attempt at recall and if the English word could not be recalled the participants were allowed to look at the other side of the card. This sequence was continued until the students felt they had control over all of the words. The second group were given cards with the English and Turkish words on the same side of the card and nothing on the other in order to exclude self-testing. This group were given a time limit for study and informed they would be tested on the words when the time was up. The third group did not have a time limit but studied under the same conditions as the second except for this condition. The second and third groups also had their packs shuffled after each study period. Following the completion of the study periods each group was tested by attempting to recall the English word after being shown the Turkish equivalent. The results showed that the self-testing group (group 1) scored an average of 19.3 of the 20 target words and that the difference between this group and group 2 (16.9 mean score) was significant. The difference between group 1 and group 3 (17.95 mean) was not significant although it should be noted that group 3 were allowed and used more time than the other groups. It is also interesting to note that group 3's self-assessment of when they had learnt the words proved to be over optimistic whereas the self-testing procedure followed by group 1 provided a more realistic measure of successful learning.

Schmitt's (2008) systematic review of the research on instructed language learning concludes that use of L1 translation in the form of word cards is an efficient method of building vocabulary size especially in the case of lower-level learners. He suggests that this is the easiest way of learning meaning and frees cognitive resources for the additional task of learning word form or spelling. Schmitt goes on to claim that currently popular pedagogic approaches lack any real focus on vocabulary and do little to promote retention; he laments the failure of published materials to clarify and implement a systematic approach to vocabulary learning arguing that the lack of recycling in course books means that partially learnt meanings are not retained.

The fields of Psychology and Psycholinguistics also offer support for the impact that retrieval and testing have on developing memory and retaining vocabulary. Roediger and Butler (2011) argue that retrieval practice is crucial wherever the aim is retention. However, they also describe a possible difficulty; if items are not correctly understood at the outset the errors are likely to be retained. This issue is further complicated by research that points to retrieval that requires greater effort leading to better long-term retention. The problem is in finding an approach which ensures correct initial understanding whilst demanding a high-level of effort. The authors suggest that Landauer and Bjork's (1978) seminal work on the spacing and organisation of retrieval attempts may provide at least part of the solution. This research found that when the intervals between retrieval attempts were gradually lengthened retention rates were higher than for equally spaced or massed, where attempts are made in quick succession. This effect appears to be provoked by the greater degree of difficulty involved in retrieval after longer periods between attempts than in immediate massed attempts or those were the interval remains constant. However, some other more recent studies have found no advantage for expanded space

retrieval over equally spaced intervals. In their review of the literature Balota, Duchek and Logan (2007) found little significant evidence for any superiority of expanded interval retrieval although they did note that there may be particular circumstances when it will produce better results. The authors call for further research citing Bjork's (1999) discussion of how the degree of difficulty in retrieval may have an impact on long-term retention. They theorise that, in longer term applied studies as opposed to shorter experimental ones, expanded interval schedules may contribute to the degree of difficulty and possibly lead to greater long-term retention.

The concept of expanded interval strategies working well when retrieval is difficult receives support from Storm, Bjork and Storm's (2010) research study. Their study involved purposely inserting a task designed to interfere with retention of the original target information between retrieval attempts. The results indicate that when information is more difficult to remember an expanded interval schedule produces better retention than an evenly spaced one. A further experimental study suggests that expanded interval retrieval may provide another benefit (Kang, Lindsay, Mozer & Pashier, 2014). The results show that although initial successful retrieval and the outcome of the final post-test were not significantly different for the expanded or equally spaced conditions the expanded interval condition produced better mean re-call over the total training period. The authors are aware of the limitations of the study in terms of length but posit that expanded interval retrieval practice may be more beneficial in areas such as language learning where study often continues over periods of years rather than weeks. They suggest that further longitudinal research would be useful. Schmidt and Bjork (1992) review a number of studies which offer support for the theory that what may appear to aid learning in the classroom in fact has a detrimental effect on long-term retention. They suggest that techniques

such as massed practice do produce immediate results but do little to aid retention. The degree of difficulty is again central to their argument with the techniques which seem to be useful in making life easier for participants paradoxically lessening the likelihood of retention. They go on to note that learning can only be seen to be successful when measured on long-term retention.

Karpicke and Roediger (2008) conducted a study comparing the effects of testing vocabulary versus studying vocabulary on retention of words once these had been initially recalled correctly. Four conditions were studied: one were study and testing of all of the target words continued; a second were testing continued on all words but once a word was recalled successfully it was not studied; a third were successfully recalled words were not tested but study of these continued; and a final condition were once words were recalled successfully they were neither studied or tested further. The results from the final test, a week after study had ceased, clearly show that testing and not studying was the key to retention. The participants in both conditions were testing was present recalled around 80% of the target words with no significant difference between the groups. The retention rates of the other two groups were 36% for the third condition and 33% for the fourth and final one. The authors report that there was no overlap in the performance between the first two and last two groups. They also claim that in many classrooms the norm is that once something is deemed to be learnt it is no longer tested although testing periodically, in other words forcing retrieval, appears to be the key to long-term retention.

In summary, this section has considered empirical and theoretical studies, from the perspectives of both Second Language Acquisition and Psychology, which have clearly demonstrated the critical role of retrieval practice in learning and retaining vocabulary. This section also focused on the related question of how the spacing of this retrieval practice might

best be organised in order to maximise benefit. Whilst there is continuing debate over the respective merits of extended and equal interval spacing there is also overwhelming agreement that both are effective and preferable to massed practice in promoting vocabulary retention. Another important factor which emerged from the literature was that of the degree of difficulty involved in retrieval. Many of the studies reviewed strongly suggest that a high degree of difficulty results in better long-term retention because the greater mental effort employed helps establish stronger memory traces. This section also claims that word cards are an excellent tool in developing vocabulary size. They can be used for self-testing at set intervals and they provide a high degree of difficulty in retrieval. Word cards, with the target word on one side and the translation on the other, also furnish a built in safeguard against the danger of an initial incorrect meaning being inferred and allow students to focus on both the meaning and form of the target language. A key limitation in many of the studies cited appears to be their short duration. Indeed, the researchers involved have often called for longer-term studies in order to further test claims of the seeming efficacy of the approaches used in promoting long-term retention (Karpicke and Roediger, 2008; Kuo and Ho, 2012).

The Importance of Recycling in Building Vocabulary Size: Breadth and Depth

The preceding section highlighted the importance of repeated retrieval practice if newly learnt vocabulary is to be retained. However, the number of meetings required to learn a word is not yet clear with studies ranging from six encounters (Saragi, Nation & Meister, 1978) to up to 20 **(**Waring & Takaki, 2003)**.** There is agreement that knowledge of a word in all aspects increases the more times it is met (Laufer, 2009; Schmitt, 2008). The frequency of meetings may be further complicated by the existing level of the language learners. Zahar, Cobb and

Spada (2001) found that very low-level learners required significantly more encounters than more advanced learners in order to learn new vocabulary. These authors also noted that two similar studies (Horst, 2000; Horst, Cobb & Meara, 1998) provided similar evidence with a high-level group (Horst, 2000) needing fewer meetings than a lower-level group (Horst, Cobb & Meara, 1998). There is almost universal agreement that to build greater knowledge of a word it must not only be encountered on several occasions but also in different ways. Table 3 outlines Nation's (2008) guidelines for teachers regarding the different ways to obtain knowledge about a word.

Table 3

Guidelines for Word Knowledge

Meaning	Form and meaning	Is the word a loan word in the L1?
	Concepts and referents	Is there an L1 word with roughly the same meaning?
	Associations	Does the word fit into the same sets as an L1 word of similar meaning?
Form	Spoken form	Can the learners repeat the word accurately if they hear it?
	Written form	Can the learners write the word correctly if they hear it?
	Word parts	Can the learners identify known affixes in the word?
Use	Grammatical functions	Does the word fit into predictable grammar patterns?
	Collocation	Does the word have the same collocations as an L1 word of similar meaning?
	Constraints on use	Does the word have the same restrictions on its use as an L1 word of similar meaning?

Source: (Nation, 2008, p.100).

It is clear that these areas cannot be effectively covered by word cards alone, yet how best to approach this issue systematically is a matter of some debate. Commercially produced coursebooks might seem to be the obvious answer: they are widely used (Richards, 2001) and cover levels from beginner to advanced whilst furnishing a ready-made syllabus (Mares, 2003). Yet on closer examination the treatment of vocabulary in a variety of coursebooks and coursebook series appears to be anything but systematic. Cobb (1995) investigated coursebooks in the context of preparation for the Cambridge PET exam and found that they provided insufficient lexical coverage and too few meetings with newly introduced words. Brown (2011) examined vocabulary treatment in nine coursebooks at three different levels and found that the overwhelming focus was on the form-meaning link with little or no attention given to the other aspects Nation identifies. In common with Cobb he also noted that there were not enough encounters with lexis to provide a strong possibility of retention. Matsuoka and Hirst (2010) analysed the New Headway Student's Book Upper-Intermediate (Soars & Soars, 2005) and concluded that this popular coursebook provided little possibility of learning vocabulary other than that in the first 2,000 most frequent words and some words from the Academic Word List. However there did appear to be better opportunities of deepening knowledge of around 30% of the 2nd 1,000 most frequent words which would be encountered five times or more if all of the coursebook were used. Given the level this book is aimed at, the treatment of vocabulary is again disappointing if we accept that frequent meetings with new words are vital. O'Loughlin's (2012) study focused on three levels (elementary, pre-intermediate and intermediate) of New English File (Oxenden &Latham-Koenig, 2006; Oxenden et. al., 2004, 2005) which is another best-selling series. This study investigated the coverage of the first 2,000 most frequent English words

based on West's General Service List (GSL), (1953). The results show that learners will only encounter around 1,500 of these words if they complete every unit of these three levels. It is also important to note that O'Loughlin's methodology counts one exposure to a word as sufficient to mark it as covered; the study does not take recycling into account. New English File's effectiveness in building vocabulary size also appears to be compromised by the amount of low frequency words which are met on only one occasion and thus are unlikely to be learnt. A further issue is the slow rate of coverage of the vocabulary that is present in the series. The author argues that for many part-time learners of English the coursebook provides the syllabus and material for a year's study, therefore these learners seem unlikely to progress beyond more than the 1,000 most frequent words in 3 years given the lack of repeated exposure to even the 2nd 1,000 most frequent.

If we accept that coursebooks, at least in their present form, are not adequate for the tasks of building vocabulary size or enhancing depth of vocabulary knowledge systematically, then we must examine the alternatives available. Nation and Ming-Tzu (1999) found that graded readers, in the format of the Oxford Bookworms series, did provide sufficient exposure to the majority of 2,410 word families in the 6-level series and that words met at the lower levels were often recycled in the higher ones. However, the authors caution that if these readers were used as a means of building vocabulary size (breadth) learners would have to read at the rate of one book per week which may well prove difficult in terms of time for many students. The role of readers in developing vocabulary depth seems less problematic; if the readers are used as a supplement to other more explicit vocabulary teaching they would afford the opportunity to meet already partially learnt words in context. The authors also identify the issue of graded readers ending at a

low vocabulary level and call for readers to be produced up to at least the 5000 word family level in order to prepare learners for unaided reading of authentic texts.

Irvine (2006) also claims that graded readers can play an important role in support of direct teaching of vocabulary. Her study found evidence that when partially learnt vocabulary was again encountered in graded readers the knowledge of each word was expanded and recall became easier. Hill (2008) conducted an extensive review of graded readers produced in the UK and the USA. He found that the quality of these books has seen steady improvement since the 1960's and argues for their adoption not just as supplementary material but as a core component of language teaching courses. Hill argues that the greatest value of readers is in the development of reading fluency and echoes Nation and Ming-Tzu in calling for an upward expansion in the vocabulary range of readers. Iwahori's (2008) study found that when students read a large number of graded readers in a seven week period their reading speed rate improved significantly. Quinn and Nation (1974) make similar claims for speed reading where learners read short texts composed of vocabulary at well below their actual vocabulary size.

In summary, this section has argued that learners need the opportunity to meet words on a number of occasions and in different contexts in order to build breadth and depth of vocabulary knowledge and that coursebooks do not meet this requirement. In contrast, it seems that graded readers may have an important part to play in deepening knowledge of already encountered vocabulary. Unfortunately, as Hill (2008) notes it is still not usual for readers to be included as a core part of language courses. It can also be argued that at very low-levels it may be more important to build size through meaning-form knowledge before useful reading can take place.

Incidental Vocabulary Learning

In contrast to the idea of an interface relationship between intentional and incidental learning (Ellis, 2005), much of what has occurred in the language classroom in the modern era has been heavily influenced by a seemingly solid belief in incidental vocabulary learning. However, a closer examination of the supporting evidence, or indeed the lack of it, seems to suggest this may have been misplaced (Raptis, 1997). For many years the 'psycholinguistic guessing game' approach to vocabulary in reading texts dominated the teaching of reading and vocabulary to L2 learners. This was how Goodman (1967) described the strategy L1 learners used when faced with new vocabulary and he later expanded his view (Goodman, 1973) to include L2 learners in a universal theory. This was seized upon and popularised in L2 teaching by writers such as Grellet (1981) and Nuttall (1982), which led to a situation where actively teaching vocabulary was not seen as important. However, there was a problem; there was little evidence to support this in L1 research and none in L2 (Grabe, 1991; Stanovich, 1980). At best it appears to be plausible to suggest that higher level learners may learn unknown vocabulary from the clues in the surrounding known vocabulary and the context of a text. However, at lower levels, where the learners have a low vocabulary size, this seems to be counter intuitive. Whilst this theory has been challenged it still has a strong influence in many classrooms and, until recently, found support from text book authors (Day & Bamford, 1998). Trainee language teachers are often encouraged to avoid translation and L2 learners to use monolingual dictionaries in preference to bilingual, although research suggests that translation could be an effective means of teaching the meaning of unknown words (Laufer & Shmueli, 1997; McKeown, 1993; Mishima, 1967).

Vocabulary Acquisition through Reading

Krashen (1985, 1989, 2003, 2004, & 2013) has consistently argued that comprehensible input through reading is all that is required to increase the vocabulary size of second language learners. Furthermore, he is constant in his argument that language is primarily acquired and not learnt. His view of acquisition is based on the 'input hypothesis' (Krashen, 1982) which posits that input containing some language just above the learner's existing level, *i + 1,* provides the best conditions for acquisition. In his earlier work he focuses on the acquisition of grammar (Krashen, 1982, 1985) and later extends this view of acquisition to include vocabulary (1989, 2004, and 2013). Pitts, White and Krashen's (1989) study focuses on two chapters of the novel 'A Clockwork Orange' with the aim of confirming that vocabulary is acquired through reading. The participants in the study (n = 74) were enrolled in intermediate level ESL classes at community colleges in California. Thirty 'nadsat' (artificial words from the novel) words were tested immediately after reading the two chapters and the results show that the two experimental groups made small gains of this vocabulary. The mean gains were 1.81/28 for Exp. group 1 and 2.42/30 for group 2, the same control group was measured against each experimental group and predictably, as they had not read the text, their gain was just over zero.

It has been highlighted that immediate post reading tests, as used in the above study, are less than satisfactory because retention is a key measure in whether words have truly been learnt (Waring & Takaki, 2003). Notwithstanding these concerns, Pitts, White and Krashen claim that despite the small gains recorded, and with no guarantee that even these were enduring, the study provides evidence of the importance of incidental learning for increased vocabulary size. They suggest therefore that second language learners 'should do a great deal of comprehensible

reading' (p.275). This recommendation seems to imply that reading is an efficient path to vocabulary growth; however, their study shows nothing of the sort. There is no comparison with any other method and therefore, this type of learning may be less efficient than others. Whilst it is not in doubt that some new vocabulary can be learnt in this manner, the authors also seem to ignore the implication that, given the rate of acquisition suggested by their study, second language learners would be required to devote unfeasible amounts of time to study if an adequate vocabulary size was to be achieved. Cobb (2008) notes that even where studies have shown reading rates at levels that would possibly allow for sufficient text to be read, the material has been graded reading at low-levels in comparison to the language levels of the participants involved. This type of reading may provide an excellent opportunity for the learners to improve reading fluency because the graded reading, below their own level, would contain vocabulary already known; for the same reason it would not provide the unknown vocabulary necessary for incidental acquisition to occur. Cobb also cautions that even in the few studies citing large volumes of extensive reading, vocabulary growth has been slow.

A number of other studies (Dupuy & Krashen, 1993; Nagy, Herman & Anderson, 1985; Pitts, White & Krashen, 1989; Sternburg, 1987; Waring & Takaki, 2003; Webb, 2008) have focused on the question of how much vocabulary can be learned incidentally through reading; however, whilst vocabulary gains have been reported most have been small. Waring & Takaki (2003) carried out a study involving 15 intermediate (or above) Japanese female students. The study used 'A Little Princess', a 400 headword graded reader, as the context for the target vocabulary to ensure that the subjects would understand the running words as the text is well below intermediate level. The spelling of the target words was altered to render these unknown

to the subjects. The authors were interested on how frequency of occurrence impacted on retention and so the 25 target words were arranged into five groups of five with different occurrence rates for each group, ranging from one occurrence for the first group to 15 to 20 for the fifth group. Three post-tests were given; one immediately after reading; another after one week; and the final test after three months. The latter two tests were unannounced. The subjects were tested for recognition of form and meaning with meaning tested in two ways; with and without prompts. The tests were administered in the same order on each occasion, with the form test first, followed by the translation test without prompts and finally by the multiple choice prompt test. The change in spelling of the words meant they would not be encountered again in the periods between tests ensuring no further opportunity for learning. The results show that after three months a mean score of less than one of the 25 target words was retained on the translation test. The form only recognition test gave a mean of 15.6 on the immediate test but this had dropped to around eight after three months. Similarly, only 6.8 words were retained on the multiple choice test at the three month stage. The authors contend that the three month translation test demonstrates that the participants had almost completely lost any knowledge of meaning and suggest that other studies, where only multiple-choice measurement was used, may have grossly inflated incidental learning gains from reading.

Incidental Acquisition through Reading: Assumptions

The claim made by many researchers that learning a large vocabulary, in a first or second language, must be connected to incidental acquisition through reading rests on a number of assumptions (Goodman 1967; Smith 1975; Sternberg 1987; Krashen 1989, 2013; Schmitt and Carter 2000). The original claim: "Incidental learning of words during reading may be the

easiest and single most powerful means of promoting large-scale vocabulary growth" (Nagy and Herman, 1987: 27) does not seem to be supported by hard empirical evidence, rather it is a default position based upon the assumption that children do not learn their L1 vocabulary through formal instruction. Therefore, it is assumed they must build their vocabulary through reading even though there is a lack of supporting research evidence. Nagy, Herman and Anderson (1985) and Nagy (1997) do not attempt to explain how this incidental acquisition ensues and acknowledge that the increased vocabulary size identified might equally be due to a number of other factors. Subsequent research indicates that children learn a substantial L1 vocabulary before they can read, although the size of the vocabulary gained varies and appears to be influenced by the socio-economic class of the child (Suggate et al., 2013). These authors also note that over focusing on the benefits of early reading fail to consider the possibility that children learn vocabulary from informal teaching encounters with parents and playmates, or that, once at school, they use the explicit vocabulary strategies they have been taught in their free time.

Notwithstanding these reservations, even if we accept the importance of reading in L1 vocabulary learning there are major differences between this and L2 vocabulary learning. A further important, and perhaps questionable, assumption is that incidental learning of vocabulary through reading comes about by learners inferring or guessing the meaning of unknown words through contextual clues. As we have seen, the L1 reader starts to read with an already substantial store of vocabulary in terms of meaning and sound; this is not the case with L2 learners at lower levels. Laufer (1992) points out that the top-down approach, borrowed from L1 acquisition theory, is unlikely to be effective with elementary level L2 learners who, by

definition, have very low vocabulary sizes. Folse (2004) goes further, describing the possibility of L2 learners benefitting from implicit learning through context clues whilst reading as a myth, brought about by possible similarities between L1 and L2 vocabulary acquisition being too readily accepted. Vocabulary learning is therefore associated with self-efficacy, defined as an individual's belief in his or her capabilities to organize and execute the courses of action required to manage a prospective situation (Bandura, 1997).

If, in the face of obstacles, the meaning of new vocabulary may be guessed, two additional requirements must be met for learning to occur; the meaning must be correctly inferred, and the inferred meaning must be retained. There is no guarantee that meaning is always correctly guessed from contextual clues (Nation, 2001). Carpenter, Sachs, Martin, Schmidt and Looft (2012) conducted a study involving eighty undergraduate students of introductory German randomly placed into four groups of twenty. Each of the groups were asked to learn 16 target words, embedded in a modified German children's story, using different approaches; these approaches included one of inferring only whilst the others included either feedback or a marginal gloss. All of the groups initially inferred a high proportion of incorrect meanings though the groups receiving corrective feedback or glosses repeated very few of these errors on intermediate and delayed tests. The inferring only group performed significantly worse with 17 of the group retaining their initial incorrect ideas of meaning on the post-tests. These results suggest that if low-level students of a second language are left to build their vocabulary through reading alone what they learn may well be counter-productive. Parry (1991) reports that her longitudinal case studies, with three strong intermediate level students and one other pre-intermediate, also revealed a high-level of incorrect or partially correct inferences. She notes that

some incorrect meanings were repeated, or near synonyms of the incorrect meaning given, in later tests where the participants supplied glosses of the target vocabulary. These tests were after the target words had been met again in different contexts and Parry points out that the danger of poor inferring may possibly have a cumulative effect.

Moving on to the second requirement, retention of correctly inferred vocabulary, we encounter another problem; the conditions which seem to aid correct inferring of meaning do not appear to help retention of that meaning. Mondria and Wit de Boer (1991) reported that pregnant contexts (where other words in a sentence give strong clues to the target word's meaning) were a useful aid to guessing word meanings and resulted in more correct guesses than from non-pregnant ones. However, they found that this improved rate of successful inference did not lead to improved retention, and noted that incorrect inferences led to better retention rates, which supports Parry's point above on the issues associated with incorrect guesses. The authors suggest this may be because pregnant contexts do not provide the degree of difficulty necessary to fix the guess in memory; the extra difficulty associated with words that are difficult to infer may achieve this aim, whether or not the meaning is guessed correctly.

Other research has cast further doubt on the usefulness of inferencing from context when the aim is to retain new or partially learnt vocabulary (Hu & Nassaji, 2012). This study involved advanced level ESL students (n = 11) reading for business/economics degrees at a Canadian university. The results indicated that when a target word was relatively easy for these students to infer from context it was unlikely to be retained and conversely, words which were difficult to infer were far more likely to be retained. Interestingly, further investigation through a think aloud protocol revealed that the words easily and correctly inferred, yet not retained, were guessed

using contextual clues only (meaning focused strategy), whilst the participants employed form focused strategies, based on word form properties, on the ones that proved more difficult to infer but were better retained. This suggests that, even with higher level learners, whilst well developed inferencing from contextual clues may be a very useful reading skill it does not appear to be an efficient method of building vocabulary size through retention of the words inferred.

These criticisms of learning vocabulary through reading are not new or restricted to second language learning. Kelly (1990) compared formal, where the guess is based on the word form, and contextual guessing. He concluded that both types were problematic although formal guesses were more likely to lead to the correct meaning. Kelly maintains that contextual guessing takes more time and that even words correctly guessed are seldom retained. Na and Nation (1985) advocate guessing as a strategy when items are low frequency and therefore unlikely to be encountered again. This is not a ringing endorsement, rather a strategy for dealing with lexis which probably does not have long term value.

Cain, Lemmon and Oakhill's (2004) studies focused on the L1 vocabulary development of 9-10 year old children. Their first study found that weaker readers were less able to guess the meaning of new words from contextual clues when the clues were not obvious; this appeared to have little effect on their better skilled counterparts. A second study was then undertaken to investigate whether the less skilled readers were either less able to infer meaning from contextual clues in particular or simply poorer at acquiring vocabulary in general. This study involved explicit teaching of vocabulary in order to compare this approach to context aided inference of meaning. The study also considered existing vocabulary knowledge in addition to reading comprehension skills and included three groups of participants; skilled readers with good

vocabulary knowledge (SRGV); less skilled readers but with equal vocabulary (LSRGV); and less skilled readers with less vocabulary knowledge (LSRLV). The first group (LSRGV) performed equally well on the inference and direct instruction tasks; the LSRGV group were less able to infer meaning than the SRGV group yet performed as well on the direct instruction task; the LSRLV group performed worse than the other groups on both tasks, requiring more repetitions in direct instruction to fix the meaning in the memory before recall was successful. However, once meaning through direct instruction was established the three groups' performance on the delayed recall test revealed no significant difference.

The authors suggest that working memory capacity may have an effect on the ability to retain meaning inferred from contextual clues. The two less skilled reading groups had lower working memory capacity than the skilled readers and both groups performed poorly on the post test of inferred meaning. The writers make no claims about the efficacy of one method over the other and, clearly, these L1 children have a larger vocabulary size than low-level L2 learners of English. Biemiller (2003) and Biemiller and Slonim (2001) claim that the average yearly vocabulary gain for native English speaking children is around 800 words up to the age of 8 and 900 between the age of 9 to 12, this would equate to a vocabulary size of between 7,300 and 8,200 for 9 to 10 year olds, yet learning vocabulary from context still appears to be more difficult than from direct instruction except for the most skilled readers.

Mason and Krashen (2004) conducted a study purportedly designed to demonstrate the greater efficiency of *listening to a story* than *of listening to a story plus form focused vocabulary activities*, when the aim is to increase vocabulary size. However, on closer examination the study (n = 58) consists of two groups explicitly studying the same 20 target words; both groups

completed a pre-test task of translating the target vocabulary into Japanese and both repeated the same task as an initial post-test. The group which was more focused on form spent a further ten minutes on oral comprehension questions before the first post-test whilst the story only group took the test immediately. The authors report that the form focused group performed only a little better on the first post-test with a mean score of 15.1of the 20 target words compared to 13.9 for the story only group. The form focused group then spent a further forty five minutes studying the target language through; retaking the translation test; checking answers with the teacher; reading the story; underlining the target words; and retelling the story to partners. This group then took the translation test once more, this time achieving a mean score of 19.7. Finally, both groups took a delayed post-test five weeks later resulting in the form focused group retaining much of what they had learnt and achieving almost double the mean score of the story only group. The authors (p.183) admit that in reality both groups are focused on form but argue these results demonstrate that the 'extra focus' on form of the 'story plus' group is not efficient as they do not learn as many words per minute. However, caution should be exercised when making such claims based on this evidence. The assumption seems to be that if these groups were subjected to further post-tests the story only group would retain this supposed efficiency advantage. It is equally plausible that, without further meetings with the target vocabulary, the story only group would in fact lose what had been gained (Grabe, 2004, Nation & Chung, 2009).

In summary, Krashen and Mason's claim that sufficient vocabulary can be acquired through reading or listening alone appears to have little merit; what seems clear is that the rate of growth if this method is followed is not remotely sufficient for the needs of second language learners and that more efficient methods of learning are required.

Arab Learners and Reading

Cobb (2007) describes how the introduction of the University of Cambridge Local Examinations Syndicate's (UCLES) Cambridge Preliminary English Test (PET) in the 1980's, led to a situation where students in Oman 'failed the test in droves, took more courses and failed it again, and too often were eventually expelled from colleges to retire to their families in some degree of disgrace' (p.13). This situation has hardly improved in the intervening years, the IELTS Annual Reviews (2006, 2007) show the participants from the UAE as having the lowest mean scores in academic reading of the 20 countries where the exam is most frequently taken; the Cambridge ESOL research notes (2010, p.29) show candidates with Arabic as a first language in the overall penultimate position for IELTS academic and bottom in IELTS general. The scores for IELTS reading paint a darker picture still, with Arabic mother tongue candidates coming last in both tests with the only mean scores below band five. It can be seen from the above that the present situation is not ideal for Arab students who wish to pursue their studies in an English medium institution. O'Sullivan (2009) underlines this point by suggesting the many Arab students do not have the reading levels necessary to achieve success in English medium education. He goes on to argue that a lack of automaticity in basic word recognition hampers their efforts to become fluent readers and to successfully use top down reading skills. Paradoxically although higher reading level skills need a solid foundation of vocabulary before they can be of use, the absence of lower level processing skills forces learners to rely on contextual clues in a situation when this is least likely to lead to comprehension (Koda, 2005). The following sections explore why Arab learners in particular appear to have chronic problems in building a solid foundation of vocabulary which can be automatically accessed when reading.

Arab Learners and Word Recognition

Whilst all second language learners face difficulties as they strive to build a usable vocabulary base, those with a first language in a different script encounter additional complications (Arab-Moghaddam & Senechal, 2001; Wang, Koda & Perfetti, 2003; Wang & Koda, 2005). Meara & Randall (1988) claim that research has focused on high-level reading skills, such as discourse organisation and text interpretation, but that little work has been done on text processing at the word level, which appears to be the core issue for those learners with scripts that are entirely different to Roman. Their study revealed that Arab learners continue to use the same U shaped search function used when reading in Arabic when reading in English; this was true of both high and lower level students in complementary experiments. This manifests itself in recognition of central letters in a word more quickly than those at either end, whereas, skilled native English readers lean towards an M shape recognising letters in the centre and at each end of a word more quickly. The study does not identify why this should be the case but the authors postulate that the underlying rationale for native Roman alphabet speakers adopting the M shape must be strong and go on to suggest that Arabic speakers will always be at a major disadvantage when reading Roman script if they are unable to adapt their strategies. The study also found that the reading speeds of the Arab learners were exceedingly slow in comparison with native English speakers and, perhaps more tellingly, very moderate compared to English language learners who were accustomed to Roman script. The authors suggest that the key to the slow reading speeds may be the lack of redundancy found in Arabic script compared to Roman script, which means Arabic is read very deliberately and that Arab learners of English may read Roman script in the same manner, leading to very slow processing of English words.

This research poses some interesting questions but is more limited in solutions; the authors suggest that further work on Arab learners and word recognition strategies is needed without giving detail of the form this might take.

The focal point of Ryan and Meara's (1991) study is also word recognition; in this case the retrieval of correct spelling. The pilot study involved sixty words of eight or nine letters with one letter in the fourth of fifth place deleted. The subjects, Arab learners, non-Arab learners of similar overall proficiency and a group of native English speakers were asked to reproduce the original word. The possible error categories were: a) no attempt; b) an attempt with incorrect spelling; and c) production of a different word. The authors report that the Arabic speakers produced at least double category c) answers compared to the other groups. Ryan and Meara attribute this to the orthography of Arabic words which commonly consist of roots of three consonants which are then combined with vowels to produce word families with the same core meaning. They point out that the major emphasis in Arabic script is on consonants and cite the example of modern written Arabic where short vowels are generally omitted. The authors compare this to written English, where similar consonant structures do not necessarily signal a semantic link and vowels assume much greater importance in conveying meaning, and conclude that this must certainly be confusing for Arab learners and could possibly explain poor performance in reading and writing Roman script. This theory is explored further in the study's main experiment which asked the three groups to consider word pairs presented on computer. The words were shown on the screen for one second then disappeared for two seconds and finally reappeared for a further period of two seconds. The word pairs were left with exactly the same spellings or, alternatively, one vowel was removed in the reappearance. The task was to

identify if the words were identical or if they had been altered and timing between each word pair was left to the participants. This experiment produced similar results to the pilot study; the Arab learners scored very poorly; as expected the native speakers achieved excellent results; and the non-Arab group performed at intermediate levels. Interestingly, when the words were unchanged on reappearance there was no significant difference between the non-Arab group and the native speaker group with both scoring highly. However, the Arab learners still produced 9.75% incorrect responses whilst their response time for this category was extremely slow, both in comparison with the other groups and with their own response times in other categories. The authors concluded that the results might well have been expected given their probable reliance on consonant roots and their tendency to neglect vowels but conceded that this is not the only possible interpretation and called for similar studies involving experiments where consonants were deleted and vowels left unchanged. They suggest that if Arab learners performed significantly better on this type of test, the data would corroborate the premise that Arab learners depend on consonantal depictions of English words and that this appears to stem from the orthography of their native language. Notwithstanding the limitations of this study, it seems clear that Arab learners have particular issues when processing English words and that solutions which focus on these particular problems are required.

Further research in a similar area explored possible connections between spelling knowledge and reading proficiency with a group of intermediate Arab students and a group of non-Arab learners of a similar level. It was found that a listening test revealed that the listening comprehension skills of the groups were comparable, with the Arab group scoring slightly but not significantly higher; in contrast the non-Arab group attained significantly higher scores on

both the reading and spelling tests. This seems to suggest that, at least in the case of Arab learners, spelling plays a more important role in reading than phonological processing. The author goes on to argue that there is a clear link between spelling and reading ability in that automatic word recognition can only develop as spelling skills develop (Fender, 2008). This view corresponds with evidence from the field of Psycholinguistics which contends that without automatic and accurate access to words, skilled reading cannot develop and that a precondition for this is well learned spelling (Perfetti and Hart, 2001; Randall, 2009). Ehri (2005) explains how well learned spelling allows the reader to recognise words as a whole (sight vocabulary, Schmitt, 2002), without having to decode through individual letters; this in turn liberates mental resources for higher level reading processes such as inferring arguments in the text.

We have seen that automatic word recognition and accurate spelling are vital to the proficient reader and that it appears that Arab learners of English are particularly deficient in these areas. It seems then that these learners need to follow strategies that include a clear focus on improving spelling, which leads to the question of what these should be. Krashen (1989) claims that spelling, in common with all acquisition, will be taken care of through reading and his input hypotheses. However, he is unable to supply any convincing evidence for this and relies on one study he was involved in (Polak & Krashen, 1988), which the authors admit provides no causal evidence of a link between reading and improved spelling, and one other (Pfau, 1967) where supplementary reading resulted in no improvement in spelling. Krashen's theories have attracted scathing criticism over the years (Gregg, 1984; McLaughlin, 1987; Mitchell & Myles, 1998) and are mentioned here solely because of their influence on classroom practice

(Thornbury, 1991; Waring & Nation, 2004), which has arguably contributed to the scarcity of explicit spelling teaching in L2 classrooms.

Widdowson's comments (as cited in Thornbury, 2018, response 20, 19.57) on Krashen's influence on L2 pedagogy perfectly capture the dangers of adopting ideas uncritically:

> My purpose in dwelling on this theory is to demonstrate how ideas are spread by the action of persuasion on uncritical acquiescence and get converted into solutions, which are assumed to be valid everywhere, like American Express traveller's cheques.

However, despite the residual effect Krashen's theories may have on contemporary classroom practitioners it seems clear that his assumptions are not supported by the majority of recent studies. Conrad (2008) claims that reading may not provide the most efficient means of improving spelling because reading does not require retrieval from memory. She proceeds to reverse Krashen's order, suggesting that improved spelling skills, through constant practice, are significantly more likely to assist in reading development. Schmitt and McCarthy (1997) also insist that language learners must make a cognitive effort if spelling is to be learnt and retained.

Birch (2007) makes a further call for explicit learning and the teaching of spelling. She, in contrast to proponents of a whole language approach, advises that beginner learners should concentrate on the phonological links between English letters and sounds in order to construct a low-level base for reading development. Whilst this seems to be good generic advice, Arab learners of English may be better served by focusing on the written forms due to their specific problems with English orthography (Randall, 2009).

Notwithstanding this advice from researchers, it appears that their pleas go largely unheeded in the modern English L2 classroom. Cook (2004) asserts that spelling is possibly the most neglected aspect in language teaching and laments the lack of published teaching materials available, whilst it seems that the vast majority of practitioners in the UAE and Oman make no systematic effort to teach spelling and, at best, only deal with errors on an ad-hoc basis (Bowen, 2008). This is not to say that attempts to tackle the problem have not been made. Pathare (2007) proposes a systematic approach to teaching Arab learners how to spell in English. Her materials are firmly grounded in recent research and are designed to focus on the particular problems Arab learners encounter.

In summary, this section has emphasized the importance of word recognition to all learners of English and examined those specific to Arab learners. It has been argued that without automaticity in word recognition learners cannot hope to become proficient readers and that this automaticity will only be developed through an explicit focus on teaching vocabulary in general including meticulous attention to spelling (Ehri, 2005; Fender, 2008; Randall, 2009). The words of Cobb (2007) sum up the present position 'the vocabulary needs of Arabic learners must be organized and planned for because they will not be met by magic' (p. 117).

Vocabulary Size Threshold, Text Coverage and Reading

A considerable number of studies have attempted to identify a threshold vocabulary size level which would provide sufficient text coverage to allow fluent reading and comprehensive understanding of a text (Carver, 1994; Hu & Nation, 2000; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Nation & Waring, 1997; Schmitt, Jiang & Grabe, 2011; Schmitt & Schmitt, 2014; Stæhr, 2008; Sutarsyah , Nation & Kennedy, 1994). Evidence of the

vocabulary size of L2 readers that is needed to read with speed and comprehension would be of appreciable benefit to researchers, course designers and practitioners alike. However, previous research findings have proved inconsistent although perhaps not contradictory. The inconsistent figures for a useful threshold seem to vary because of differences in estimates of the coverage required to read with understanding. During the past 30 years much more information has become available on the coverage needed to facilitate the ease of reading a variety of text types. As the knowledge base in this area has grown, coverage levels have been consistently revised upwards with a corresponding increase in the vocabulary size L2 learners of English must attain to meet these coverage levels. Thus, it seems apparent that the crucial variable is how much of the lexis in a text must be available to the reader in order to understand it (Schmitt, Jiang & Grabe, 2011).

Laufer's (1989) study found that 95% coverage of an academic reading comprehension exam text allowed the majority of the participants at this threshold to reach a comprehension level of 55% in a multiple-choice test. Laufer calculated the in-text vocabulary known by each student by asking the participants to report which words they did not know and through the results of a translation test. These findings highlight a further important issue; at which point can comprehension be described as adequate? As the subjects in this study were potential university undergraduates 55% text comprehension appears to be a surprisingly low requirement. It also seems likely that were the comprehension requirement raised, greater vocabulary coverage would be needed to meet the higher target. In addition to the limitation of the comprehension requirement, Laufer used the 95% coverage level to calculate the 5000 word vocabulary size required based on Dutch language frequency count research (Ostyn and Godin, 1985). This

raises two further objections; the text types used in the Dutch research were a combination of newspaper articles with academic texts and the question of whether research on Dutch language frequency counts is relevant to studies on the English language (Nation, 2001). The first objection is founded on several studies which have revealed that the vocabulary size required for high text coverage varies depending on the genre to be read (Hirsh, 1992; Nation, 2006; Nation & Waring, 1997; Sutarsyah, Nation & Kennedy, 1994). Despite these limitations, Laufer's study has provided a platform for more recent work in this area by suggesting a linear relationship between comprehension, coverage and vocabulary size. Additionally, as we shall see later, her conclusions, given the relatively low-level of comprehension sought, correlate well with later research.

The varying vocabulary size demands to achieve high-levels of coverage in different genres are demonstrated by Hirsh and Nation (1992). Their study investigated the vocabulary size needed to read a novel written for English teenagers. This text type was chosen in the belief that the vocabulary used for younger readers might be less sophisticated than that in a novel aimed at adults and, secondly, that a relatively long text focusing on one theme and written by a single author might result in more frequent repetition of the lexis (Nation and Waring, 1997). The results showed that for this genre a vocabulary size of 5,000 words achieved 98.5% coverage whilst 2,600 words provided 96% coverage. The higher figure should allow good comprehension of the text and even 96% coverage would appear to make around a 60% comprehension score possible if we relate this study to Laufer's above. Nation (2006) acknowledges that the Hirst and Nation study suffered from the lack of suitable frequency lists then available and notes that later developments may now allow similar studies to provide

superior guidance on vocabulary size and text coverage. Similar findings were reported by Sutarsyah, Nation and Kennedy (1994) in their comparison of the amount of word families found in a single large economics text (5,438) and the number found in a text of approximately the same size, consisting of a variety of different academic genres (12,744). These studies indicate that texts on particular subject areas or on a specific topic place far less demands on vocabulary size. The implications for pedagogy include the possibilities that such texts may be useful for recycling mid-frequency vocabulary (Schmitt and Schmitt, 2014) and that concentrating on the teaching of specific high frequency subject vocabulary could provide an improved vocabulary size to text coverage ratio in the field of English for Academic Purposes (Coxhead, 2000; Ward, 1999).

Hu and Nation (2000) conducted a similar study to that of Laufer (1989) above, in a further attempt to link vocabulary coverage with reading comprehension. This study measured adequate comprehension by means of a multiple choice test, in common with Laufer, and through a written recall test. In the interest of comparing the results of the two studies only the multiple choice test results are considered here. As we have seen Laufer's study set the adequate comprehension bar at 55% comprehension which, as the author admits, seems remarkably low. In contrast, Hu and Nation opted for a seemingly more reasonable adequate comprehension score of 12 correct responses from 14 or approximately 85.7% comprehension. Their original text only contained words from the 2,000 most frequent English words; they then used 'nonsense' words to manipulate the original text and create four coverage levels of 80%, 90%, 95% and 100%. The results showed that no participant achieved adequate comprehension at 80% coverage; a few reached the required level at 90% and 95% whilst nearly all did at 100% coverage. The authors

concluded that 98% coverage is needed for adequate understanding. However, Nation (2001) notes that although the two studies suggest different coverage levels this does not cast doubt on either due to the difference in where the adequate comprehension level is set. He comments that if we seek adequate comprehension, here meaning almost everything is understood, for a high majority of learners then 98% coverage is the likely threshold, although if a lower level of understanding at Laufer's 55% comprehension level is sought then 95% coverage is probably sufficient. These two studies also indicate that as vocabulary size grows, reading comprehension improves and that the necessary coverage level of any text depends on the degree of comprehension sought. The 98% coverage finding tallies with that of Carver (1994); he found that native speakers read effortlessly when all words were known, but that when only 2% of words were unknown reading became relatively more difficult.

Nation (2006) revisited the question of what vocabulary size was needed for comprehension of the majority of texts, this time with the aid of the British National Corpus (BNC) frequency lists. He advocates a greater focus on targeted vocabulary learning, especially for lower level learners, in order to build a vocabulary size of around 8000 to 9000 word families which he suggests would provide around 98% coverage of authentic reading texts. This percentage equates to one unknown word in 50 running words, which Nation claims would allow adequate although not effortless comprehension. As we can see a pattern now begins to emerge; whilst it is plausible that some meaning will be understood with text coverage of less than 98%, this will be achieved with great difficulty and will not be adequate for learners such as those studying in English medium environments.

Staer (2008) also investigates the links between vocabulary size and skills in English as a foreign language. The participants in the study were 15 to 16 year old Danish students (n=88) taking a Danish national school leaving exam. These students had almost completed lower secondary school, where the Danish syllabus provides for a minimum of 570 hours of English language learning over a seven year period. The instruments employed in Staer's study of the reading skill were a multiple choice and matching reading test on a range of text types, which aimed to test sub-skills including: reading for gist; detailed understanding; extracting specific information; and inferencing from the text. Vocabulary size was measured with the revised version of the Vocabulary Levels test (VLT) at the 2,000, 3,000, 5,000 and 10,000 levels (Schmitt, Schmitt and Clapham, 2001). The VLT revealed that only 20 of the 88 participants had reached the 2,000 word level which appears to have had a profound impact on reading ability. Evidence for this impact was provided by the result of a Spearman correlation between vocabulary size and reading skills which found a highly significant correlation of .83 whilst a binary regression analysis indicated that 72% of the variance in reading results was attributable to vocabulary size. The results of the reading test also reveal a mediocre average level of comprehension of around 56% which may point to low text coverage level as in Laufer's (1989) study above. The author, shocked by the results, calls for a more explicit approach to vocabulary teaching in Danish schools in place of the reliance on incidental vocabulary learning in vogue at the time of the study.

Laufer and Ravenhorst-Kalovski (2010) revisit the question of lexical coverage with a more comprehensive study which attempts build on previous studies (Hu & Nation, 2000; Laufer, 1989; Nation, 2006) by exploring the relationship between lexical coverage, vocabulary

size and the reading comprehension of academic English. The authors also address the vexing question of what is 'adequate comprehension' by providing definitions of two possible levels of adequacy. The number of participants (745) allowed the formation of groups at five different vocabulary size levels as measured by the revised VLT (Schmitt, Schmitt & Clapham, 2001). The vocabulary size that would be needed to achieve coverage of the texts used in the reading tests was calculated using a vocabulary profiler freely available on Cobb's Compleat Lexical Tutor website (n.d.). Personal and geographical names in the texts were assumed to be known on the basis that they did not belong to a particular language and the percentage of these in each text was also calculated. Thus, if a text was found to require a vocabulary size of 4,000 words to reach around 94% coverage and it contained 2% of proper names the coverage was adjusted to 96%. The participants' reading skills were tested using the English reading comprehension section of Israel's national university entrance test, which had proven reliable and had been standardised over a period of over 20 years. This test was designed to measure various aspects of comprehension including; lexical understanding; syntax; inferring; and extracting specific information. The data gathered on text coverage of texts, the participants' vocabulary sizes and their reading comprehension results was then used to investigate the relationships between these three variables. As expected, those students with higher vocabulary sizes generally performed better on the reading comprehension test; linear regression revealed that 64% of the divergence in reading scores could be credited to vocabulary size. However, a particularly interesting pattern becomes apparent when the relationship between the extra coverage gained for each 1,000 words learnt and the resultant gain in reading score is considered. Whilst common sense would suggest that learning 1000 words which provided the greatest increase in coverage would be likely to

give the largest gain in reading comprehension this did not prove to be the case. Learning the 2nd 1,000 of the most frequent words provided an increase in coverage of around 9% and a corresponding mean gain of seven points in the reading test; learning the next 1,000 most frequent words afforded almost 3% more coverage but an increase in mean reading scores of over 12%. This pattern continues when the 4th and 5th thousand words are added to vocabulary size; these increase coverage by approximately 2.2% and 1.2% respectively yet mean reading scores for both of these relatively small gains in coverage rise by around 10 points. These results have important implications for pedagogy; whereas previous research seemed to indicate little return in terms of time taken and coverage gained for the explicit teaching of vocabulary outside of the 3,000 most frequent words, it now appears that even small increases of coverage can facilitate major improvements reading skills. The authors suggest two possible reasons for the results, firstly that lower frequency words may at times be crucial to understanding and, secondly, that as vocabulary size grows, the automaticity of access to vocabulary learnt earlier improves the fluency and speed of reading. The question of the level of vocabulary size and text coverage needed to attain an adequate reading level of academic texts is answered on two levels; 4 to 5,000 words to achieve 95.5% coverage and 6 to 8,000 with 98% coverage. The writers claim that the lower level is sufficient for around 70% understanding but that the learners would require further support, whilst the higher level should enable independent reading. It seems questionable that the lower level would equip readers of academic texts for university study, where surely the ability to read independently is a prerequisite given the other competing demands made on L2 learners in English medium education. It should be noted that a similar criticism was made on Laufer's (1989) study above. These findings though do correlate with

upward revisions of the vocabulary size, to around 8,000 word families, needed to read independently and achieve 98% coverage of most texts (Nation, 2006).

Schmitt, Jiang and Grabe's (2011) large-scale study revisits the issues researched in Laufer's and Hu and Nation's earlier work. The aim is again to investigate the correlation between vocabulary coverage of a text and the understanding of that text. A yes/no checklist test, where participants' mark the words they know, was developed to test receptive form-meaning knowledge of vocabulary in the two reading texts used to test comprehension and thus arrive at a reliable measure of the participants' (upper-intermediate and above) text coverage levels. The reading texts were profiled for vocabulary frequency to identify words at the 0 to 500, 500 to 1,000, 1,000 to 2,000 and above 2,000 levels. It was assumed, given their level, the participants would prove to have above 90% understanding of the first 1,000 words. The vocabulary at the two other levels was tested in far greater depth with more than 50% of words at these levels and which occurred in either text included in the checklist. This resulted in 120 words for inclusion and in addition the test featured 30 non-words making a total of 150 words arranged in ten groups of 15 in frequency order; the non-words were randomly inserted into the groups with each group containing one to three. These non-words had been included to guard against the participants' overestimating their knowledge; if too many non-words were found to be known the participant was excluded from the study. The two reading texts were chosen to represent authentic examples of academic texts and because participants were thought to have strong background knowledge of the subject in one text but less of the other. Comprehensive reading tests were developed to test understanding of the texts. The results focus on the relationship between reading comprehension and text coverage ranging from 90% to 100% at 1% intervals.

The writers report a linear correlation with each 1% increase bringing a stable corresponding gain in understanding within a range of 50% comprehension at 90% coverage to 75% comprehension at 100% coverage. As the authors acknowledge, this study was with advanced students and it is likely that a study with lower level participants may well produce different results if authentic materials were used but only because coverage levels would be lower. Also, it is apparent that even 100% coverage of the vocabulary in a text does not lead to full comprehension because, as the authors note, other factors are involved. However, it does seem clear that, as the other studies considered in this section also suggest, greater vocabulary size leads to greater comprehension and that vocabulary size is the crucial element in reading comprehension. The writers agree with Nation (2006) that to attain a comprehension level appropriate for academic study 98% coverage and a vocabulary size of 8,000 to 9,000 word families is needed.

As you will have noted, the research in this area now points to a far higher vocabulary size requirement if learners are to achieve the text coverage needed to read a variety of texts independently and with comprehension that may be adequate for academic study in English. The studies considered above reveal a high level of agreement between researchers who have built upon and replicated the previous work of their peers and it seems apparent this advance in knowledge has profound implications for pedagogy. If we accept the stipulation of 8,000 to 9,000 word families as adequate for independent reading (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt, Jiang & Grabe, 2011) together with Laufer and Ravenhorst-Kalovski's claim that although learning each 1,000 word level from the 3rd to 7th most frequent adds little to coverage, yet provides far better comprehension, we have constructed a powerful case for an

explicit focus on these word families. This theme is taken up by Schmitt and Schmitt (2014), who argue strongly for a re-classification of what have been considered the boundaries between high and low-frequency vocabulary. They contend that the families up to the 3,000 level should be labelled as high-frequency and those from this level to 10,000 as mid-frequency, with families above this mark still named low-frequency. The authors recognise the importance of knowing this mid-frequency vocabulary and call for further research on how it might be taught. They, in common with Laufer and Nation (2001), claim that the processing speed of high-frequency vocabulary improves markedly once learners have achieved a vocabulary size of 5,000 to 6,000 words thus highlighting a reason why this vocabulary would benefit L2 learners studying in English.

In summary, this section has considered both theoretical and empirical studies on possible vocabulary thresholds, vocabulary size, text coverage and reading comprehension. We have seen that the vocabulary size thought necessary for a high level of understanding has risen from earlier estimates of 3.000 to 5,000 word families to around 9,000 families and that even this higher figure is probably the minimum when learners are reading for academic purposes. This increased learning burden has placed as yet unanswered demands on pedagogy, notwithstanding that the leading researchers in this area are in agreement on the revised figures. There is now also widespread consensus that around 98% lexical coverage of a text is needed to allow reasonably fluent reading although there are important caveats to take into account. The first of these is that a lower level of comprehension is possible with lower coverage; studies indicate that 95% coverage should allow between 50 and 60% understanding. Secondly, even the 98% figure in isolation does not afford total understanding; other factors influence reading comprehension

although vocabulary knowledge and lexical coverage are agreed to be by far the most important. The review of the literature in this section also reveals a consistent acceptance that growth in vocabulary size not only facilitates improved comprehension but also contributes to gains in reading speed as automatic access to more lexis becomes available. This seems to identify vocabulary size as being of particular importance for English learners studying in English medium environments. The research does not confirm a specific vocabulary threshold which promotes an immediate and significant increase in comprehension, nevertheless, it does demonstrate that all growth in vocabulary size, even after the most frequent words are learnt, leads to corresponding improvements in reading skills.

Vocabulary Learning Strategies and Self-Efficacy Theory

The above sections outline the present position of Arab learners of English in terms of vocabulary and reading levels. This position can be summarised as one where failure has become the norm; students are asked to attempt reading texts that are beyond their compass given their limited understanding of English vocabulary. In the case of low-level Arab learners, their ability to increase vocabulary size rapidly is arguably compromised by using learning strategies that are unlikely to lead to success and which in turn reinforce their lack of belief in their own abilities.

Vocabulary learning is therefore associated with self-efficacy, defined as an individual's belief in his or her capabilities to organize and execute the courses of action required to manage prospective situation. Self-efficacy theory is a component of social-cognitive theory, positing that in order to achieve any specific outcome individuals must first believe that they possess the tools to facilitate that outcome (Bandura, 1997). Self-efficacy theory is considered by many researchers in education to be a critically important contribution to the study of academic

achievement, motivation, and learning, and this theory also provides a foundation for the development of educational policies and classroom practices (Artino, 2012).

By personally helping students to improve their individual mind-sets, teachers may help individual students to improve their perseverance, learning behaviour, and academic achievement. Hamada (2014) found, using qualitative data, that some students with a poor mind-set lost their enthusiasm for learning English. Hamada identified five strategies to improve mind-set and prevent demotivation of English language learners, of which the teachers' sensitivity was found to be the most important strategy. Teacher sensitivity included items that highlighted active engagement with the student, respect for the learners' individuality, support for efforts and activities, including noticing and praising improvement. These motivating factors were not directly related to specific techniques for teaching English, but rather they reflected the teachers' positive attitudes toward the students. In the context of the present study, it is argued that these strategies were present but directly related to the word card treatment in the case of the experimental group. As the students worked in pairs using the cards the teacher was able to monitor and provide help whenever needed. The teachers of both control and experimental groups supported students in their learning and gave positive feedback wherever possible. The class sizes were capped at a level allowing the teachers to interact with individuals, pairs and small groups. In this regard, there is much empirical evidence in the literature to support the hypothesis that smaller class sizes can lead to more active engagement of the teacher with individual students, leading to enhanced individual academic achievement (Chingos & Whitehurst, 2011; Ehrenberg, Brewer, Gamoran, & Willms, 2001).

Rauber and Gill (2004), conducting qualitative research in Brazil, found that a diversity of teacher's responses (e.g., correcting, agreeing, appreciating) enabled students to participate in the classroom in order to extend their English language proficiency. The ability of a teacher to apply a diversity of responses, intuitively and spontaneously, and to ensure that they were understood, was appreciated by students and resulted in higher levels of motivation.

Online resources and vocabulary learning

There is continuing debate as to the effectiveness of web-based learning. One major criticism raises concerns over the low rates of interaction between students, their peers and their teachers in online environments (Arbaugh, 2000). Another related issue highlights feelings of isolation contributing to students either not completing online tasks of not devoting sufficient time to those tasks that are completed (Tyler-Smith, 2006). Macaro, Handley and Walter (2012) argue that much of the impetus for the use of Computer Assisted Language Learning (CALL) in English language learning has come from governments and their educational authorities with the aim of increasing digital literacy in populations. They call for further research to investigate and provide evidence on where and why CALL might be useful from the perspective of educational theory. An in-depth analysis of CALL is beyond the scope of this study. However, the issues above were kept in mind when considering how CALL could be ulilised as part of the experimental treatment in the present study.

One positive aspect of online resources appears to be simply that students enjoy using them and also appear to benefit from their use. Research in Malaysia found that 300 undergraduates believed that social interaction platforms could support their English language learning. The writers called for further research into how these platforms could be blended into

structured learning programmes in order to target specific language skills (Kabilan, Ahmad, & Abidin, 2010). Almekhlafi, (2006) study of the use of Computer-Assisted Language Learning (CALL) with year 12 male high school students in the UAE, found that the students in the experimental CALL using group were strongly in favour of its use and believed that their English skills had improved as a result. The quantitative results from the study supported the students' belief as the experimental group's language learning performance was significantly better than that of the control group, which did not use CALL. The researcher noted that the results also suggested that those students with superior computer skills performed best. He concluded that training in computer skills should be given in advance of the implementation of CALL programmes.

Other aspects of research in this area include Chien's (2015) qualitative study of three web-based vocabulary flashcard sites. The main aims were firstly to evaluate the websites in terms of the options they provided for teachers and students and, secondly, to investigate how students felt about using flashcards and online sites for vocabulary meaning. The study took place in Taiwan with intermediate first year undergraduate students (n=64). Data was gathered through observations of classes when students were using the vocabulary sites and focus group interviews with 20 of the participants in the study. Chien found that the students were generally positive about the approach to vocabulary learning and felt it had aided their success in broadening their knowledge. The classroom observations highlighted the importance of ensuring that students knew how to use each site before learning began. This was identified as a key element of the teachers' role. The study concluded that all of the sites focussed mainly on the form-meaning link between individual words. A criticism was that no features were available to

develop depth of word knowledge. Further criticism centred on the lack of options for timing or recording students' performance. The issue of early clarification of how to use online software was also identified in a study at Northern Illinois University involving 20 students studying Filipino. The researchers found that time spent on explicit clarification before and in the first one or two weeks of a programme saves time overall. This study also recommended the development of tracking options so that in future studies researchers could compare the participants' reported strategies with the actual strategies they used (Gallo-Crail & Zerwekh, 2002).

A claim made in favour of the use of vocabulary games in vocabulary learning is that games can kindle students' interest in learning and help maintain motivation. Yip & Kwan's (2006) mixed method study in Hong Kong focused on comparing the performance of control and experimental groups. The experimental group were recommended to use two websites selected by the researchers and involving vocabulary games, whilst the control group would attempt to learn the same vocabulary through classroom based activities. The participants (n=100) were all first year Engineering undergraduates. The experimental group were introduced to the workings of the website at the beginning of the study and then asked to focus on learning the designated words through the websites (in class) with minimal further teacher support. The teachers' role then became one of monitoring the students to ensure they were only working on the prescribed sites. The results showed a significant difference in the performance of the two groups. The experimental group had clearly learnt the vocabulary better than the control group. The qualitative data revealed some interesting results. The majority of the experimental group (68%) was found to favour online learning over activity based learning. The students found that the recycling in the online environment helped them retain meaning. The focus group interviews

provided more sophisticated insights to the students' perceptions. They opined that the competitive nature of some of the games increased their motivation as they sought to increase their game scores. However, they also cautioned that some of the vocabulary scores may well have been higher if they had not been driven to increase their game speeds and not over focused on the game in detriment to the learning. The survey results showed that one of the vocabulary game sites received more positive feedback than the other. The interviews clarified that this was because the site which received the less positive feedback posed the students with more difficulty in understanding the game instructions.

In summary, the key point from this section is that online vocabulary learning is not per se more efficient or motivating than classroom based learning. Online learning is more motivating when the instructions to students are clear; in contrast, positivity over a game is diminished when they are not. Efficiency of learning can lessen if the game itself becomes more important than the learning. It also seems evident that, in order to achieve maximum benefit from online learning, students should be taught computer skills as a pre-requisite. Observations from a pedagogical viewpoint stress the need for online activities to include options for recording students' performance so that learning effects can be tracked and measured. In addition, online learning should be closely monitored to ensure that the positive aspects of interaction with peers and teachers are not lost. Finally, as with all teaching materials, online learning must be informed by evidence.

Translation

The use of translation in language teaching and learning has been, at best, discouraged and at times prescribed for long periods of the past 120 years. We might ask why this should be. There are number of possible explanations, most of which appear to be for convenience, profit or outdated theory rather than current evidence. The first of these is the common assumption that when a new language is learnt it should be learnt monolingually, that is with little or no reference to the learner's own language. This theory is particularly convenient in the case of teaching English, where many teachers around the world are young, with little or no training and without knowledge of any language but their own. Secondly, it is more profitable for educational publishers to be able to produce a single product, which can be marketed globally, than to produce course books aimed at national markets. In addition, monolingual teaching fitted well with early British and American SLA theory and with the much promoted communicative approach. Any suggestion that a learner's own language might be useful when learning English posed uncomfortable challenges to the preceding positions (Auerbach, 1993; Hall & Cook, 2012; Phillipson, 1992; Ramachandran, Devi & Rahim, 2004).

Although there has been some discussion of and movement towards the use of the learner's own language in recent years, the use of translation remains taboo to many in the profession (Cook, 2010). There are relatively few studies focusing on language learners' beliefs about translation as a vocabulary learning strategy. However, the available studies show that, in contrast to the teaching profession, many language learners seem to have faith in the value of translation. Liao's (2006) study focused on the beliefs of 351 Taiwanese students on the use of translation when learning English. The researcher used three questionnaires to gather data; one

collected background information, another investigated beliefs about translation and the third sought data on how often translation was actually used as a learning strategy. The questionnaires were followed up with focus group interviews with two groups of five students from the higher and lower levels of English proficiency within the sample. The author reports that "the participants overwhelmingly believe that translating helps them acquire English language skills…" (Liao, 2006, p.201), including vocabulary. This sample included students majoring in English (51%), who were able to articulate their thoughts on the advantages and disadvantages of translation, and students with other majors who were less proficient in English. Interestingly, the author notes that various levels of English proficiency made no significant difference to the participants' beliefs about translation or its use as a strategy. Schmitt (1997) also conducted a survey in his study of 600 Japanese participants at four different age levels. He set out to identify the strategies which were most used and to elicit which were found to be most useful. As with Liao's study, translation, here in the form of bilingual dictionaries, proved extremely popular with language learners, ranking first in both categories.

In the context of this study, it would be useful to review authoritative studies on Arab learners' beliefs on translation. Unfortunately, studies published in leading journals are rare. However, local studies exist and are reported below. A survey of 124 low-level English foundation students at a UAE university sector institution revealed that the students were generally in favour of the use of Arabic to support English language learning. The author notes that Arabic was officially banned in university level English classes but that the ban was politically rather than pedagogically motivated. The motivation being founded on the poor results from secondary school English classes, where teachers have a poor level of English and

classes are often carried out exclusively in Arabic (Mouhanna, 2009). Two points can be made in this case: The overuse of Arabic in UAE secondary schools is not a result of teachers' beliefs but rather their lack of adequate English; this has clearly resulted in a reaction officially ruling out even the judicious use of L1 at university level.

Machaal's (2012) study at a Saudi college surveyed 197 male students and interviewed 13 teachers and 3 members of senior management. The students were all in their college foundation stage with ages ranging from late teens to early twenties and with, as the author reports, low attainment in English. Overall, the survey data shows that 63% of the students supported the use of Arabic in their classrooms. However, it should be noted that 86% of the students felt that Arabic should be used to facilitate their understanding of new vocabulary. The majority of the teachers interviewed (77%) expressed a belief that Arabic could be used as long as there was a clear aim and rationale for its use. This is in a context where teachers were aware of and had been trained to use a communicative approach with little place for Arabic in their classrooms. Additionally, even those teachers who stated opposition to the use of Arabic conceded it might be useful when teaching vocabulary. Further studies found that in Oman L1 was used by many teachers because an English only classroom was very difficult to implement with low-level learners (Al-Hinai, 2006). Al-Alawi (2008) found a split between those teachers in favour of its use and those who were not. Finally, another study in Oman revealed tension between practitioners who followed a communicative approach and their students who wanted more use of Arabic (Al- Shidhani, 2009).

Guessing from Context

The theoretical arguments in support of a 'guessing' approach to unknown vocabulary were discussed above. Major disadvantages with this strategy include the strong possibility that learners will guess an incorrect meaning; that it gives a low return in relation to the time taken; that in order to successfully guess an unknown word in a text a learner would need to already know a 95-98% of the other words (Laufer, 1997; Nation, 2013). However, even in light of the above, this strategy has been advocated by many theorists, teacher trainers and teachers.

Incidental Learning through Reading

We have seen above the arguments that the only strategy needed to build vocabulary size is to read as much as possible (Krashen, 1985, 1989, 2003, 2004, 2013; Pitts, White & Krashen, 1989). The objections to this strategy are many. Firstly, it is extremely difficult for low-level learners to read in any meaningful manner without first learning vocabulary. Secondly, the amount of reading required would be so large as to render any gains painfully slow (Cobb, 2008). In addition, the lack of measurable progress from the student's viewpoint is unlikely to increase motivation (Bandura, 1997).

Learning Vocabulary from Graded Reading texts

Graded readers are books that have been explicitly written for children or for second language learners. These readers are either adapted from existing work, including classic works of literature, film scripts and travel books, or original work written specifically for the genre. In both cases these texts are simplified to allow language learners easier access to the story. The simplification focusses on grammar and vocabulary. In the case of grammar, sentence structure is often changed from complex to simple and extra signposting, in the form of cohesive devices,

added (Crossley, Allen & McNamara, 2012). Vocabulary is graded in terms of the frequency of the headwords used. For example, a graded reader series might limit level 1 to texts containing only the first 400 headwords in a frequency list and then increase the amount of headwords permitted in subsequent levels.

The crucial question in the context of this study is whether graded readers can significantly increase the vocabulary size of elementary level learners of English in a reasonable time frame. The answer appears to be no. Studies have suggested that the learners would need to know a minimum of 95% of the running words in a text in order to read with any fluency and to have the opportunity to incidentally acquire any unknown meanings (Nation & Ming-Tzu, 1999). The participants in this study knew very little English vocabulary at the outset and would have been challenged by even the simplest of graded readers. Additionally, the reading load indicated for useful acquisition is very high. Recommendations include one graded reader a week and 30 minutes reading a day (Waring & Nation, 2004; Yamamoto, 2011) for relatively small gains. When the lack of a reading culture in of elementary Arab learners of English, even in their L1, is taken into account, the possible efficacy of extensive reading of graded readers, as a means of significantly increasing vocabulary size, seems extremely doubtful. The disadvantage of a lack of reading culture can be seen to be magnified in the case of male Emirati students by the roles they are expected to play in life outside of college. These include escorting female relatives when they are in public and taking an active role wider family life (Harb & El-Shaarawi, 2007). This leaves little time for reading once formal study time has ended.

However, graded reading may have a very important role to play in vocabulary development. This role appears to be in increasing the depth of knowledge on already known

vocabulary and aiding retention of that vocabulary (Waring and Takaki, 2003). This would suggest that graders reading might be most beneficial once students have learnt the core meanings of the most frequent words and can read texts that have a vocabulary load within their current vocabulary size.

Explicit Contextualised Learning through Coursebooks

Advocates of contextualised learning have argued that strategies must involve initial learning in context if full understanding of vocabulary is to be gained (Gardner, 2007; Webb, 2008). However, whilst this approach may be useful with advanced learners, it is less likely to be a successful strategy for learners at an elementary level, due to the need for reaching a vocabulary threshold before learners can understand the context (Laufer, 2003; Nation, 2003). Put simply, at elementary level there are likely to be too many unknown words.

Coursebook writers have attempted to solve this problem by including a focus on the most frequent vocabulary in their work. However, issues such as the amount of recycling once a word is met and the slow rate of progress if a coursebook is followed do not appear to have been addressed (Brown, 2011;O'Loughlin, 2012).

This section has outlined potential vocabulary strategies and highlighted possible strengths and weaknesses. It has also raised the point that no strategy is likely to be successful unless the students believe in that strategy.

Summary

This chapter has reviewed literature central to the position of vocabulary in language learning and of the relationship between vocabulary and reading, both of which are of critical

importance to the present study. Firstly, the historical lack of focus on the learning and teaching of vocabulary in the main language teaching methods and approaches was highlighted. Subsequently, the key concepts in this study of the nature of word knowledge and the construct of reading ability were discussed. This was followed by the crucial question, in the context of this study, of whether vocabulary should be taught explicitly or acquired implicitly. The definitions in this research context were defined. A further issue, that of if vocabulary is better learnt in or out of context, was also considered in this section. The next section focused research studies on decontextualised vocabulary learning and included a particular focus on word cards and translation which is the experimental treatment in this study. The section brings together studies from linguistics, SLA and psycholinguistics which provide complementary evidence of the benefits of retrieval in vocabulary learning. Finally, this section investigates the importance of recycling of vocabulary through learning materials.

A review of the theory of incidental vocabulary learning follows and includes research studies on acquisition through reading and listening. The assumptions of acquiring vocabulary through reading are examined. The chapter then moves on to investigate the problems faced by Arab learners when reading in English. Next the issue of the specific text coverage provided by various vocabulary sizes is explored. This issue is of major importance to the present study given the reported low vocabulary sizes of many Arab learners of English and their problems when reading in English. Finally, a selection of strategies for vocabulary learning is outlined.

Vocabulary size and its relation to reading performance have attracted steadily increasing attention from researchers in the past 30 years. The claim that sufficient vocabulary for independent reading can be acquired incidentally through reading has been strongly criticised.

Empirical evidence has suggested that an explicit focus on vocabulary learning is required if vocabulary size is to grow at rate which would allow elementary learners to quickly reach a threshold adequate for independent reading. Factors such as low vocabulary size, word recognition and student beliefs have been found to have a bearing on the poor reading performance of Arab learners of English.

Whilst these issues have been examined it should be pointed out that much of what the research has to say has not been applied to classroom teaching and learning (Folse, 2010; Matsuoka & Hirsch, 2010; Nation, 2003; Schmitt & Schmitt, 2014). In particular, there is little detailed longitudinal research focusing on the vocabulary size and reading ability of elementary level Emirati students of English. Mixed method research, using quantitative and qualitative methods is even more difficult to find. A number of researchers in the UAE have highlighted the issues facing students in the foundation years of English medium higher education institutions and have called for further classroom based studies, which aim to discover the most efficient approaches to building vocabulary size (Davidson, Atkinson & Spring, 2011; Watts, 2011). In addition, given recent research (Schmitt, 2008), there appears to be a need to investigate the effect of translation, which has long be frowned upon in the UAE, as a tool in vocabulary learning. Furthermore, the beliefs of Emirati students in these institutions should be sought. These students have generally not been successful and may have developed a belief that what is demanded of them in terms of vocabulary and reading cannot be achieved.

Conducting this review has allowed me to identify some of the key issues in vocabulary learning in general, how reading ability can be developed and, in each area, those issues which

appear to be specific to Arab learners of English. These issues and how they have been addressed in my research design and methodology are discussed below.

The issue of whether vocabulary is best learnt incidentally or explicitly has been a central discussion point in the literature. The possible use of word cards as an effective tool in a programme of explicit vocabulary learning has been posited by many researchers. However, few, if any, studies have been carried out in authentic classroom setting over a full academic year. The experimental word card treatment in this study is designed to provide controlled practice of the most frequent English vocabulary. The reading materials used by the experimental groups are purposely set at a vocabulary level below the students' existing vocabulary size in an effort to ensure that the experimental students would not learn vocabulary incidentally through their reading.

Multiple studies from the fields of SLA, Psychology and Psycholinguistics concerning the strong effect of retrieval in vocabulary learning were reviewed. The overwhelming consensus supported the view that learning materials which forced students to retrieve from memory created strong links with the target vocabulary and promoted automaticity of recall. Word cards, in contrast to other approaches such as word lists, force students to retrieve from memory because the students cannot see the target meaning. The treatment in this study was planned so that retrieval was spaced. This planning ensured that the students would continue to meet and retrieve meaning over a prolonged period.

The researcher also reviewed the literature on the importance of recycling in order to build both breadth and depth in vocabulary learning. Whilst there was dispute over the number of encounters needed to learn a word, there was consensus on two points of importance to this

study. Firstly, that knowledge of a word deepens the more times it is met and, secondly, lower level learners, as in the present study, need considerably more meetings than higher level learners, if they are to learn and retain new vocabulary. The studies reviewed cast doubt on whether it was likely that the amount of recycling required could be provided through the use of conventional classroom materials such as coursebooks. This study was designed to address these issues by ensuring the experimental groups encountered their target words a minimum of sixteen times through word cards and online practice.

As this study incorporates the use of online recycling of vocabulary, this relatively modern approach to learning was reviewed in order to identify any issues particular to online learning. These issues were taken into account in the design and use of online material in the experimental treatment. In addition, the vocabulary met in the reading material used by the experimental students (speed reading texts; graded readers; adapted texts profiled for vocabulary load) ensured that these students would again encounter the words, initially met through the decontextualised treatment, now in context when reading.

The literature review also revealed that although the use of translation in language teaching has been actively discouraged for much of the past century, the reasons for this do not seem to be based on any compelling evidence, especially in the case of vocabulary learning. In contrast to what might be termed the orthodox view on the use of translation in language teaching, the literature suggests that many language learners, including Arab learners of English, supported the use of translation. This support was even more pronounced in the case of vocabulary learning. As translation is at the heart of the experimental treatment in the present study, it seemed appropriate to investigate the views of the students involved. This element was

incorporated into the mixed method research design through survey questionnaires and focus group interviews. The survey questionnaires and focus group interviews also explored the students' beliefs about vocabulary learning strategies and reading skills in the light of self-efficacy theory. This theory holds that students are only likely to learn successfully if they believe the desired learning outcomes are attainable and that they possess the necessary tools and strategies.

The final major findings from the literature review concern the problems specific to Arab speakers when reading and learning vocabulary in English and how these are addressed in this study. The literature highlights the poor performance of Arab candidates in international reading tests and suggests that their inability to recognise basic word forms is at least partly to blame. This deficiency appears to be linked to the strikingly different ways readers of Arabic and readers of English process texts. The studies reviewed strongly recommend that a specific focus on English spelling is needed as Arab learners of English build working vocabularies. It is suggested that this would improve their ability to recognise whole words on sight and lay the foundation for developing higher level reading skills. The field of Psycholinguistics advises that, without this automaticity of word recognition, vital mental resources are wasted on low level processing, leaving few resources for higher level skills. This study explicitly focuses on learning the spelling of the most frequent English words through the use of word cards. Each time the cards were used the students were required to give the English translation of the Arabic word and spell out the English word. This was reinforced by online phonological processing when the students listened to the words and then wrote them.

This study attempts to fill the gaps in knowledge identified above through investigating an experimental treatment of vocabulary teaching and by capturing the students' beliefs on the best strategies for learning vocabulary. Two key research questions have therefore been articulated below to investigate if a translation based approach to vocabulary learning is a successful strategy and, if so, what, if any, effect does increased vocabulary size have on reading ability.

CHAPTER 3: Methodology

Introduction

Methodology in social science refers to the approach of the researcher towards explaining and understanding the social world (Babbie, 2010). This chapter begins with an introduction to the researcher's methodology by restating the research questions that were addressed using a mixed methods approach, which for reasons outlined in Chapter 1, were syntactically orientated so that they could not be answered by either "Yes" or "No".

RQ1: To what extent, and in what ways, does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element?

RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores?

RQ3: What are the perceptions of elementary level Emirati learners of English regarding the learning of vocabulary and its relationship to reading comprehension?

The two research hypotheses tested in this study were as follows:

H1: Decontextualised vocabulary study, using word cards and translation, may contribute toward a more rapid gain and a greater amount of receptive vocabulary among elementary level Emirati learners of English, than a similar teaching programme lacking this element.

H2: The receptive vocabulary size of the elementary Emirati learners correlate with the PET reading scores. In contrast, the IELTS reading scores correlate only with the receptive vocabulary size of those participants who exhibited the greatest receptive vocabulary gains.

The researcher did not adopt the traditional method of testing null hypotheses vs. alternative/research hypotheses. For nearly a century, researchers have believed that statistical significance, based on the computation of inferential test statistics and associated p-values, should be interpreted as reliable evidence to reject null hypotheses and/or accept alternative/research hypotheses (Huberty, 1999). The main reason for not using null hypothesis tests was that the absence of statistical significance at an arbitrary level (e.g., $p > .05$) does not provide reliable evidence to prove that a null hypothesis is true. As Alderson, (2004) argues 'Absence of evidence is not evidence for absence' (p. 328). It is not possible using inferential statistics to prove definitively the existence of nothingness (Sorensen, 2015). Therefore, if the results of this study did not support the research hypotheses, then the absence of statistical evidence did not imply that the stated research hypotheses were false. Lack of statistical significance implied only that insufficient data was provided, and that judgement must be suspended until new data becomes available in the future (Hurlbert & Lombardi, 2009 Furthermore, the rejection of a null hypothesis using an arbitrary level of statistical significance , conventionally $p < .05$ would not provide reliable evidence to prove that the research hypotheses were true (Chia. 1997; Filho, Paranos, da Rocha, Batista, Silva, & Santos, 2013; Haller, & Krauss, 2002; Halsey, Curran-Everettt, Vowler, & Drummond, 2015; Hubbard & Lindsay, 2008; Huck, 2009; Nuzzo, 2014; Wasserstein & Lazar, 2016). '... surely, God loves the .06 nearly as much as the .05' (Rosnow & Rosenthal 1989, p. 1276).

Significant objective evidence in the literature lends support to the view that relying solely on null hypothesis significance testing may be a flawed method, and should not be uncritically applied. The main criticism is that p-values alone are unreliable and irreproducible

measures of statistical evidence, which are a function of the sample size. It seems questionable whether it can be proved or disproved definitively that a hypothesis is true or false using the conventional $p < .05$ criterion. Hurlbert and Lombard (2009) explain that Sir Ronald Fisher, the developer of ANOVA in the 1920's, did not propose a threshold p-value to reject a null hypothesis (e.g., there is no significant difference between mean values if $p > .05$). Fisher originally suggested that the p-value is a very informal and subjective measure of statistical evidence to interpret the differences between mean values. However, Fisher's suggestion was not widely applied in practice. Fisher's view was challenged in the 1920's and 30's by Neyman and Pearson (1928; 1933) who formulated a new decision-making framework for inferential statistical tests. This framework required an all or nothing dichotomous decision to be made between a "significant: or a "not significant" outcome depending on whether the p-value was less than or greater than .05. The Neyman-Pearson theoretical framework has dominated statistical inference since then and became part of the institution of statistical testing. However, there is objective evidence to demonstrate that the theory underpinning this framework has been weakened, promoting Orlitzky (2012) to recommend that null hypothesis significance tests should be 'deinstitutionalised'. In the light of this advice this study gave more weight to the effect sizes rather than the p values to interpret the results of ANOVA. Effect sizes do not depend on the sample size and are considered by many writers to be a better indication of the effect of an intervention in experimental studies (Ferguson, 2009; Fraenkel & Wallen, 2011; Hill & Thompson, 2004; Kotrlik & Williams, 2004). In contrast to the p-value, which takes into account the size of the sample and the size of the effect, reporting the effect size tells us the how large or important the effect of the intervention was. The effect size focuses on the mean

difference between two groups, in the case of this study the difference the intervention made to the means of the experimental and control groups.

The arguments above have more recently been supported by the American Statistical Association which issued a serious of formal statements criticising the widespread misuse of p-values and advocating a ban on the uncritical use of p-values and null hypothesis testing (Wasserstein & Lazar, 2016; Wasserstein, Schirm, & Lazar, 2019). These statements include:

> P-values do not measure the probability that the studied hypothesis is true. Scientific conclusions or policy decisions should not be based only on whether a p-value passes a specific threshold. The p-value, or statistical significance, does not measure the size of an effect or the importance of a result. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis. (2016, p.132)

Consequently, even if $p > .05$ in this study did support the research hypotheses, it was still necessary to corroborate the results using more reliable information, such as effect sizes, and qualitative data,

The researcher supported the many arguments in the literature proposing that *p*-values, which are unstable functions of the sample size, do not distinguish between important vs. unimportant results, and are the least useful and inconsistent outcomes of statistical tests. Effect sizes, that are not dependent on the sample size, appear to be more stable and useful than *p*-values to evaluate the impact of interventions in experimental research designs (Ferguson, 2009; Fraenkel & Wallen, 2011; Hill & Thompson, 2004; Kotrlik & Williams, 2004; Lipsey, Puzio. Yun, Hebert, Steinka-Fry, et al., 2012; Vacha-Haase, 2001).

The subsequent sections of this chapter consider the following methodological details:

(1) Population and Participants; (2) Overview of Mixed Methods Approach; (3) Sequential Explanatory Research Design; (4) Phase 1 Survey Methodology; (5) Phase 1: Experimental Methodology (RQ1 and H1); (6) Procedure (7) Instruments; (8) Statistical Methodology for Experimental Data; (9) Phase 1: Modelling Methodology (RQ2 and H2); (10) Limitations of Phase 1; (11) Phase 2: Qualitative Methodology; (12) Phase 3: Triangulation ; (13) Limitations of Phase 2; (14) Limitations of Phase 3; (15) Ethical Considerations.

Population and Participants

The target population for this mixed methods study consisted of foundation students entering the Higher Colleges of Technology, Al Ain Men's College (AAMC) in Al Ain, UAE. All the students were male Emiratis, with the same L1 (Gulf Arabic) background. The majority had studied English as a Foreign Language as a school subject for eight to twelve years; a few students had attended institutions such as the Institute of Applied Technology where some subjects are taught in English. The students were classified according to the scores awarded for the Common Educational Professional Assessment (CEPA) outlined in Table 4. The participants in the study were 80 students from 2 levels in the initial sample, although during the course of the study the data on 3 participants from the experimental and 3 from the control was discarded leaving a total of 74; the reasons were lack of attendance or drop out.

The participants in this study were randomly assigned to the control and experimental groups once they had been accepted by the institution. The institutional measure for entry was the CEPA examination described in the instruments section of this study. CEPA was designed to test general levels of English but does include vocabulary questions. The researcher had no control over which students were granted entry to the institution's foundation programme.

Unfortunately, it was discovered that the vocabulary sizes of the control and experimental groups were not equivalent only after the first VST. It was then too late to change due to institutional regulations. This issue is discussed further in the limitations section of the 'discussion' chapter.

The researcher does not claim that the sample is representative of all male Arab learners of English, or representative of the entire population of 11000 foundation students in the UAE, of which about 70% were women. However, it is argued in the discussion chapter that this study may well have relevance for a wider population, including all Arab learners of English in terms of issues concerning the different orthographies of Arabic and English. Additionally, the findings on the effectiveness of the word card treatment are likely to inform language teaching pedagogy for wider populations of low level language learners. Therefore, although the findings of this study may lack some external validity in terms of its ability to generalise the results from the sample to other populations, the implications for other populations should be followed up with further studies.

In order to form the groups the students from the foundation level 2 (CEPA 151-160) were randomly allocated to four groups of twenty and then two groups from this level were randomly assigned to the study. Similarly, the students from foundation level 3 (CEPA 161-170 were randomly allocated to two groups of twenty and two groups of 15 and the two groups of twenty were assigned to the study. This study was concerned with elementary level students and considered the participants from both foundation level 2 and level 3 to be at elementary level at the beginning of the study. It may be argued that level 2 is lower elementary and level 3 higher elementary, however, there was overlap between the levels. The study investigates the relative performance of the two levels, as well as the overall performance, to ascertain if level 2 was

likely to progress quickly enough to enter the college post foundation within 4 semesters, the maximum allowed. It should be noted that foundation level 2 in semester 1 of the study became level 3 in semester 2 and level 3 in semester 1 became level 4 in semester 2 in accordance with the institution's procedures.

Table 4

HCT Foundations CEPA Scores on Entry 2010

Foundation Level	CEPA Score	Number of Students	CEPA Descriptor
1	140-150	130	Extremely low English ability (Beginner)
2	151-160	80	Low, At-risk
3	161-170	70	Emerging Proficiency
4	171-180	0	Intermediate
Direct Entry	181-210	0	Ready for direct entry to English-medium tertiary study

Source: HCT 2010

The CEPA is a set of locally-developed standardised tests produced by the UAE Ministry of Higher Education used for admissions and placement by the three federal institutions of higher education in the UAE. The CEPA was originally designed as a placement test for students after entry to the foundation programme of the UAE's federal university sector. This role changed to that of a gatekeeper for entry to the institutions in 2006, and was used as such in the present study. The total time limit for the examination, which consisted of Grammar and Vocabulary, Reading and Writing papers, was 2 hours. There were no set times for each component although recommendations of 45, 45 and 30 minutes respectively have been made (Coombe & Davidson, 2014).

Overview of Mixed Method Approach

The terms qualitative and quantitative apply to different methodological approaches.

These terms not only define the types of data collected and the methods used to collect the data, but also the inquiry paradigm or philosophical perspective of the researcher (Creswell, 2014). Quantitative research is underpinned by positivism (i.e., believing that facts and feelings are separate, and that knowledge is an objective reality that exists outside the human mind); whilst qualitative research is underpinned by constructivism (i.e., believing that facts and feelings are not separate, and that knowledge is a socially constructed reality). Quantitative research methodologies, underpinned by positivism, generate numerical data that may be generalised to a defined group or population of individuals in terms of descriptive statistics (e.g., mean and variance) and inferential statistics (e.g., analysis of variance). It is, however, difficult to explain the attitudes and perceptions of each individual person through the statistical analysis of quantitative data alone. A common error of quantitative educational researchers is the ecological fallacy, defined as the assumption that each individual in a group behaves in exactly the same way as the mean value or other statistic computed to summarise the behaviour of the whole group (May, Boe, & Boruch, 2003). Qualitative research methodologies, underpinned by constructivism, address this difficulty by exploring the richness, depth, and complexity of the perceptions and attitudes of each unique individual. Qualitative research does not attempt to provide information that can be generalised to a group of people, but it may provide information that can be used to enrich or confirm quantitative data, and/or to generalise the qualitative data to a theory (Bogdan & Biklen, 2006; Merriam, 2014).

A mixed method approach is defined as the collection and analysis of both quantitative and qualitative data in parallel, or in series. The value of the mixed methods approach is that it does not accept the polarisation between quantitative and qualitative methodologies (Tashakkori

& Teddlie, 2003). Creswell & Plano Clark (2007, p. 5) highlighted the benefits of a mixed methods approach stating that: 'Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone'. Dornyei (2007) suggested that mixed method research allows the researcher to view a phenomenon from different angles thus providing a richer picture. Neither methodology is considered to be superior to the other. The researcher is empowered to use whatever methods are considered necessary to work best in practice.

Mixed methods research designs have emerged as a challenger to the polarisation of the positivist and constructivist paradigms (Symonds & Gorard, 2008; 2010). However, Hesse-Biber and Leavy (2006) argued that there may be a temptation to use mixed methods in the hope that more equals better. However, more does not always equal better because quantitative and qualitative methods are underpinned by different assumptions and philosophical perspectives that may act as barriers to the integration and collective interpretation of the data (Bryman, 2007; Creamer, 2018). The reasons for performing a mixed methods study must be made clear at the start, otherwise this approach cannot be justified. Onwuegbuzie and Teddlie (2003, p. 379) recommended that 'researchers undertaking mixed methods techniques should seek to defend explicitly the approaches they are employing'. The rationale for undertaking mixed methods techniques in educational research was emphasised by Fraenkel and Wallen (2011) stating: 'We believe that educational research increasingly is, and should be, a mixture of quantitative and qualitative approaches' (p. 430) and that 'as far as we are concerned, research in education should ask a variety of questions, move in a variety of directions, encompass a variety of methodologies, and use a variety of tools' (p. 14). It is essential to combine the analysis of

qualitative data (e.g., the responses of students to interview questions) with the analysis of quantitative data (e.g., numerical scores obtained using a questionnaire survey and/or student test/examination scores) in order to answer questions in educational research concerning the attitudes, perceptions and performance of students (Johnson & Christensen, 2004). Therefore, the mixed methods approach was justified in answering the stated research questions and associated hypotheses that guided the current study.

Sequential Explanatory Research Design

Several different types of mixed method research design have been defined (Creswell, 2014). A mixed methods sequential explanatory design was applied in the current study. The essential feature of this design is that quantitative and qualitative data are collected, analysed, and interpreted in three consecutive phases. The advantage of this design is that it facilitates a richer explanation and understanding of the research topic with more detailed answers to the research questions than could be obtained by using quantitative or qualitative data alone or separately (Ivankova, Creswell, & Stick, 2006; Bowen, Rose, & Pilkington, 2017). The implementation of a sequential explanatory design provided the researcher with an opportunity to better explain the factors associated with the assessment of the vocabulary growth and reading development of the elementary level Emirati learners of English. This design facilitated the enrichment of the quantitative assessment data with qualitative data describing the participants' attitudes and preferences regarding the teaching and testing methods employed. Figure.1 presents a diagram to outline the three phases of the research design.

In Phase 1, quantitative data were collected from samples of Emirati learners of English, using a survey and an experiment. The survey involved the administration of two questionnaires

using items with 5-point scales designed to evaluate (a) how the learners felt about different ways of learning English vocabulary; and (b) what the learners believed were their main problems when reading English. One questionnaire with 13 items (see Appendix A) was administered to the control group of learners, who were not exposed to decontextualised vocabulary study with word cards and translation. The other survey (see Appendix B) with 14 items was administered to the experimental group of learners, who were exposed to decontextualised vocabulary study with word cards and translation. The difference between the two questionnaires was the addition of one item for the experimental group concerning: 'Learning vocabulary from word cards with English on one side and the Arabic translation on the other'. With the exception of that item, the survey responses to the 13 items collected from the control group were compared with the survey responses of the 13 items collected from the experimental group.

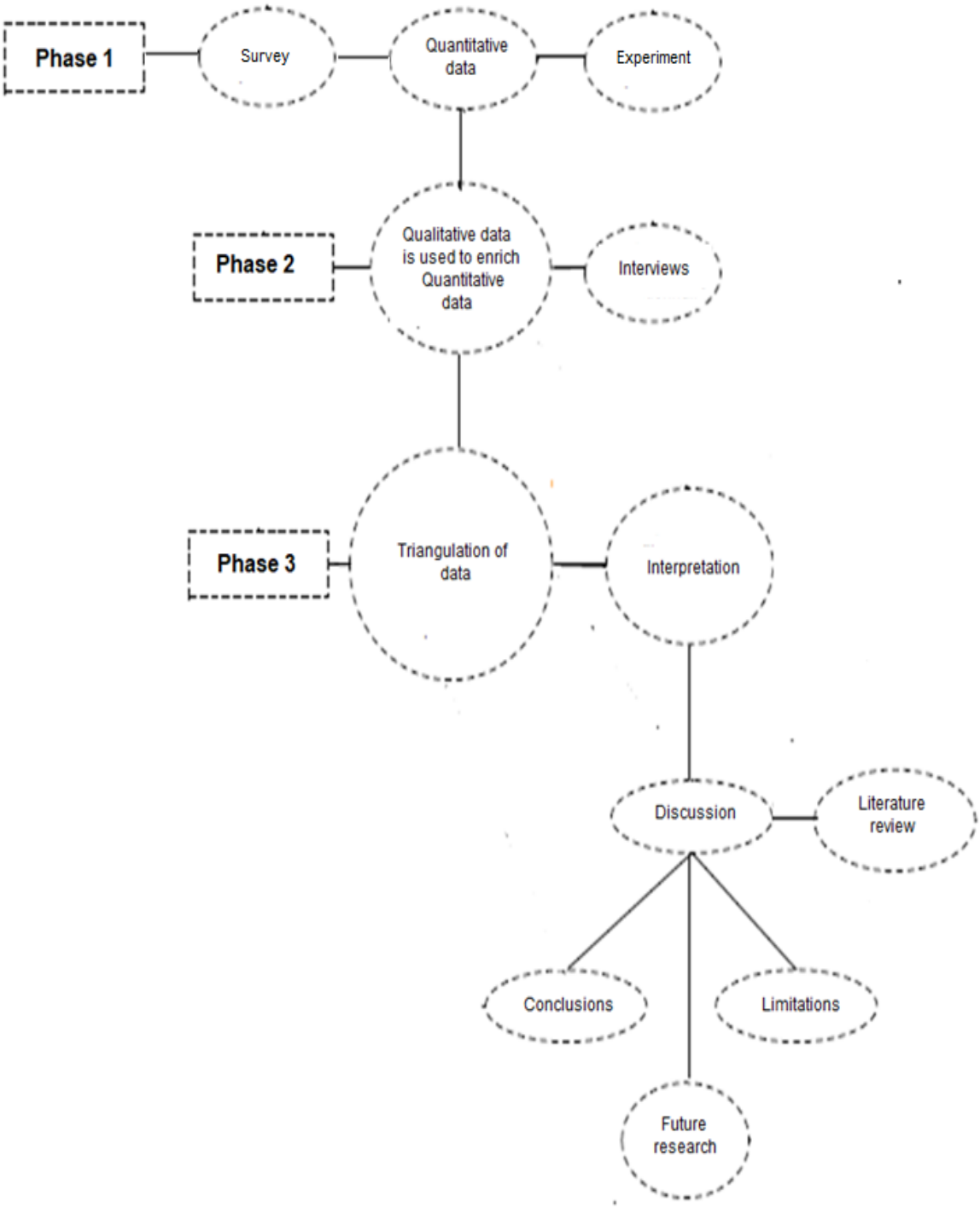

Figure 1 Outline of sequential explanatory mixed methods research design

The experiment conducted in Phase 1 involved a pretest with a multiple posttest research design. The experiment incorporated an intervention to determine the effects of decontextualised vocabulary study on the vocabulary growth and reading development of the Emirati learners of English. The examination scores collected to assess vocabulary and reading ability from the control group and the experimental group were compared. The quantitative data were analysed using descriptive, inferential and modelling statistics.

In Phase 2, the responses to semi-structured interviews were collected in order to enrich the quantitative data. The interviews involved asking a sample of learners to explore their own experiences of the examinations, the different ways of learning English vocabulary and the difficulties they experienced when reading English. A thematic analysis of the interview transcripts was conducted. In Phase 3, triangulation was conducted, in which the qualitative data and quantitative data were interpreted collectively. The final part of Phase 3 was a discussion of the results in the context of the literature, including the drawing of conclusions, a consideration of the limitations of the findings and recommendations for future research.

Phase 1: Survey Methodology

This survey questionnaire was created for this study and had been piloted with a group of students (n=12) from the same institution during the previous academic year. The pilot group had a similar English level to the main sample. These students were firstly provided with the survey instructions during a meeting with the researcher. The aim of this stage was to check the students understanding of what was required of them. The pilot participants did not report any significant issues with understanding the statements provided to them or of what they were required to do.

Internal consistency proved to be strong, with the Cronbach alpha coefficient .811 for the 9 question experimental group version and .820 for the 8 question control group version.

The pilot participants completed questionnaires to capture their views on the merits of a variety of vocabulary learning strategies. They were asked to respond on a 5-point scale, ranging from 'not useful at all' to extremely useful', to indicate how useful they had found the various strategies. The participants accessed the questionnaire online through the SurveyMonkey programme (http://www.surveymonkey.com). The questions were carefully checked using a vocabulary profiler to ensure that the vocabulary used was within the participants' range; 96.8% came from the 2000 most frequent English words or proper nouns, such as 'vocabulary, which were well known to the participants. An Arabic translation and a translator were also available to provide further clarification but were not required.

In the main study the questionnaires included items measured with 5-point scales (see Appendix A and B) were administered to a sample consisting of N = 106 Emirati learners of English drawn from the target population. The sample was divided into the control group (n = 71) who were not exposed to the intervention, and the experimental group *(n* = 35) who were exposed to the intervention. The sample size of the control group was increased by surveying the remaining students from the groups at levels 2 and 3 on entry, who were not in the experimental and control groups assigned to the study. These students had followed the same programme as the other control students. The 5-point scale, devised to evaluate how the learners felt about different ways of learning English vocabulary, was scored numerically as follows: 1 = Not useful at all; 2 = Slightly useful; 3 = Useful; 4 = Very useful; 5 = Extremely useful. The 5-point scale, devised to evaluate what the learners believed were their main problems when reading

English, was scored numerically as follows: 1 =Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree.

The researcher reviewed the literature to choose the most appropriate statistical methods to analyse the survey data. For over 50 years, there has been considerable debate in the literature regarding the most appropriate methods used to analyse questionnaire item scores based on rating scales (Carifio & Perla, 2007; 2008; Jamieson, 2004; Norman, 2010). Some researchers have argued that rating scales constitute ordinal level categories that should be analysed using non-parametric descriptive statistics (e.g., median and mode) and categorical methods of inferential analysis that do not assume normal distributions (e.g., Chi Square test, Mann-Whitney U test, Spearman's rank correlation, and Kruskal-Wallis test). Other researchers have argued that rating scales are continuous level variables that can be analysed using parametric statistics (e.g., mean, standard deviation) and parametric methods of inferential analysis (e.g., t-test, ANOVA, Pearson's correlation, and linear regression) that assume normality.

Carifio & Perla's (2008) view on this debate recommended that parametric methods were appropriate for the statistical analysis of rating scales. Their review concluded that many studies, conducted during the last 50 years, have revealed that it is a fallacy to believe that rating scales are ordinal measures that can only be analysed using non- parametric statistics. The rating scale response format produces interval or scale level data at that is appropriate for parametric descriptive analysis (e.g., mean, standard deviation, etc.). Furthermore, inferential test statistics, including the *t*-test statistic and the ANOVA *F* statistic are robust and provide unbiased results for the analysis of group differences in a rating scale, even if the scale deviates from normality.

Phase 1: Experimental Methodology

A sample consisting of $N = 80$ (74 remained at the conclusion of the study) students in foundation levels 2 and 3 was randomly selected from the target population to take part in the experiment. Their mean age was just above 19 years. Most were in the 18 to 21 age range, four were 17 years old and the others were company sponsored students between 22 and 28 years old. Foundations levels 2 and 3 were chosen because they best represented the elementary Emirati earners of English required to address the research questions. Level 1 included extremely low English ability (beginners) whose language ability was too low. The different work plans for high and low-levels in the experimental and control groups are summarised in Appendix C. Although the different work plans were important, the key issue was that the time spent on each aspect of learning vocabulary and reading was closely controlled. The experimental and control groups spent the same amount of time on each aspect of the course. The key differences between the control group and experimental group were in the vocabulary and reading treatments; the other areas of the work plans were exactly the same. The reading material for the experimental group was controlled so that they would not acquire vocabulary from their reading but instead focus on reading fluency and deepening knowledge of vocabulary they had already learnt through the vocabulary treatment. In contrast, the control group read material above their current vocabulary level with the hope that they would acquire some vocabulary through reading. The control group also worked on specialist vocabulary course books throughout the year.

To address RQ1 and test H1 the learners were assigned into either a control group who were not exposed to the intervention ($n = 40$) or an experimental group who were exposed to the intervention ($n = 40$). Equal numbers of students in foundation level 2 (n = 20) and level 3 ($n =$

20) were assigned to the control group. Equal numbers of students in foundation level 2 (n = 20) and Level 3 (n = 20) were assigned to the experimental group and the control group.

The researcher taught the experimental group whilst the control group was taught by two well qualified and experienced colleagues. Pre-course planning meetings were held with the aim of plotting what would be taught in terms of vocabulary and reading during the course. These pre-course meetings were supplemented by regular weekly meetings of the three tutors during the course. The experimental and control groups spent 3 classroom hours on direct vocabulary study and a further 3 hours on developing reading skills per week. Students were also set a further 2 hours of homework per week in each of these areas.

The Procedure: Materials and Programmes used

The experimental groups were taught using the following approach:

Vocabulary

Word cards were created in packs of twenty and were based Bauman and Culligan's (1995) updated version of West's (1953) GSL. They covered the first 2,284 most frequent words and the 570 word Academic word list (Coxhead, 2000). The cards are double sided and have English on one side with the standard Arabic translation on the other. The cards were in class sets of nine packs for each set of twenty words, allowing up to eighteen students in pairs to work at the same time. The key element of this approach is the forced retrieval of meaning and spelling each time a word is met. The critical role of retrieval in developing automaticity of word recognition is identified in the literature review (p. 34). The focus on spelling is designed to address the issues caused by the differences in orthography between Arabic and English as identified in the literature.

Learner training took place at the beginning of the study. The students were taught to work in pairs. Student A held the pack in use with the Arabic translation facing student B be who sat opposite student A. Student B gave the English target word. If Student B did not immediately say the correct target word student A turned the card to enable student B to focus on the form and spelling of the target word. The card was then returned to the back of the pack. If student B did identify the target word correctly, student A then asked for the spelling of the word. If student B was again correct, the target English word was revealed and then removed from the pack. If student B could not retrieve the correct spelling, student A revealed the English word and student B studied the spelling for a few seconds. In this case, the card was then returned to the pack. Once a pack had been completed, student A again tested student B on the cards remaining in the pack. After each practice attempt was concluded the remaining cards in the pack were shuffled in order to guard against list effects (Nakata, 2008; Nation & Webb, 2011). After completion of this stage the roles were reversed with student B now testing student A. The class tutor monitored the pairs throughout the process to ensure that the procedure was followed. The tutor also provided an initial model of the pronunciation and was available to answer any questions the students posed. Students practised until all the words in a set of 20 were known and then moved onto the next pack.

The students began using the word cards at a level just below that indicated by their vocabulary size scores. For example, if their VST score was 700 words the student would use cards from 600 onwards in order to ensure that the high frequency words were learnt as well as possible. The rationale here is based upon the premise that the acquisition of high frequency

vocabulary is very important for low level learners and that initially learning through cards would provide comprehensible input of the core meaning.

This approach was applied throughout the academic year and was reinforced by online work, which followed once the students could successfully retrieve the meaning and spell each word in each set of word cards.

Online recycling exercises

Nation (2001) discusses what teachers need to teach and what students need to know about a word, which he describes as the 'learning burden'. Table 1 of this study provides an overview of the receptive and productive elements of word knowledge. With these demands in mind, it was considered imperative that the vocabulary was initially met and retrieved on a high number of occasions in order to reinforce the form-meaning link of the words and begin to develop depth of word knowledge.

To this end two online programmes were utilised:
'Words to Know' (WTK) developed at Al Ain Women's College, and a further programme developed at Dubai Men's College (DMC). Both of these programmes follow the GSL with WTK covering the first 2,000 words and the DMC programme the first 1,000. One week after a word card pack had been successfully learnt the students began the DMC exercises on the same words online in order to provide extra meetings with the vocabulary, deepen word knowledge and to provide spaced retrieval practice. The research studies discussed in chapter 2 underline the important role of spaced retrieval practice in vocabulary learning. The online exercises facilitated the spacing of practice in the following manner. Two weeks after a word card pack and one week after the DMC online practice the students met the words again through the WTK programme

again with the aim of spacing retrieval practice. These exercises covered meaning, spelling, pronunciation and use for every set of words. Both programmes employ translation exercises. On the DMC programme meaning, spelling, listening and were tested and a score was recorded for each area before a final multiple choice test:

- Meaning, a multiple choice where learners chose the Arabic word with the same meaning as the English one.
- Spelling, a multiple choice, learners heard the English word and chose the correct spelling.
- Spelling, heard the word and typed in the spelling.
- Using, a multiple choice sentence completion.
- Final test, multiple choice. Learners chose the English word which best matched the Arabic example.

If a learner achieved less than 80% in any area they repeated the exercises and retook the test.

The teacher followed progress on the Blackboard Vista (BBV) site where the scores were automatically recorded for each student. The teacher had control over when students could see each set of words, ensuring that students followed the prescribed procedures before moving on.

The WTK course had a series of game style matching activities for each set of words and recorded the time spent on each activity before it was completed successfully. The activities are (in order):

- jig sound, where they heard the English word and matched the Arabic written form to it.
- Listen match, as above except that players match the English sound to the written Arabic form.

- Speed word, students saw the Arabic word and had to spell the English equivalent using the online keyboard. They could choose from 3 speeds, with the time to spell reduced at each level of difficulty.
- Listen spell, as above except students heard the English word and then spelt it.
- Spelling web, Players saw the Arabic word and mapped the letter paths to spell out the words in English.

Some of these games were played in pairs, in order to introduce an element of competition in an attempt to foster motivation as discussed in the online resources section of the Literature Review. There were 5 activities for each set of words culminating in a final test for each set, here the students were shown the Arabic word and had to type in the English word. Again, if the students scored below 80% they repeated the activities. The teacher was able to monitor what each student had scored online. The final test provided an opportunity to expand the spacing of retrieval. Literature in the review suggests that expanded spacing makes retrieval more difficult and that successful retrieval under difficult conditions leads to greater automaticity of word recognition.

The AWL became the focus for experimental level 3 when they became a level 4 group in the second semester of the experiment. Eleven weeks were spent working on the AWL with word cards and recycling using an online programme. As this programme could not be tracked, the work was done in class time.

The rationale for this work is founded on the belief that high frequency words must be learnt before successful reading development can take place. The AWL was important because at the end of level 4 the students were entered for the Academic IELTS test where receptive

knowledge of frequent academic words is vital for the reading exam. This recycling was designed to build on the foundation of the word cards by ensuring that learners met the newly learnt words frequently, approximately between 16 and 20 times not including incidental meetings through their reading programme. The crucial significance of recycling if word knowledge is to be retained was identified in the literature (p.39). This condition was met through both word cards and online recycling in the experimental treatment.

Control Groups

As noted earlier the foundation programmes at the HCT had been completely re-vamped for the academic year 2010/11. It might be argued that, due to this, the control groups were also taking part in an experimental procedure, although this new programme was designated as system wide. This meant that all students of the same level would follow the same courses and use the same materials for each programme strand throughout the HCT. The control groups' work plan can be found in appendix C.

Vocabulary

The control groups studied vocabulary intentionally through the Oxford Word Skills course (Gairns & Redman, 2008). This introduces vocabulary in context and is organised around topics or themes, however, there is no translation element. The level 2 control group followed the OWS Basic whilst the level 3 control group covered the final units of the basic course and the first half of OWS Intermediate. These books are based on the 'Oxford 3000' wordlist, which contains the 'the 3000 most important words to learn in English' according to the Oxford 3000 webpage (2019). This list is founded on the British National Corpus frequency lists, with account also taken of the range covered by the words included and has considerable overlap with the

GSL. The level 2 students began their study at a level above their then current vocabulary size with a starting point at around the 400 word level, similarly the level 3 group starting point at the 600 word level was set slightly above their then current vocabulary size. The initial vocabulary size test gave a mean size of around 300 words for level 2 and 500 for level 3. It is important to note that the plan for control groups, in common with the experimental students, was to focus on learning the most frequent English vocabulary. That is the same words were targeted but the approach to learning taken was different. The control groups studied explicitly in context and were also expected to acquire vocabulary through their reading.

The question of how many times the learners would need to meet a word before they fully acquired it for receptive purposes is of importance in this study. The OWS recycles the targeted vocabulary two to three times through each book whereas the experimental groups would meet each new word between 18 to 25 times, if meetings through their reading are included.

Control group work on the AWL also began at level 4 and was based on the recommended course book, Inside Reading 1: the academic word list in context (Burgmeier & Zimmerman, 2007). As the title suggests the control groups met the AWL vocabulary in context and without translation.

Reading: Experimental groups Extensive Reading – graded readers

As the learners developed their vocabulary size, as described above, they also spent class time on using class sets of graded readers at the rate of one book round every 8 days or 10 per semester. The readers used were selected to be at or just below their actual vocabulary size. The books were chosen from two well-known series, Oxford Bookworms and Penguin readers, in

order to get full coverage of vocabulary levels. Table below shows how the readers compare with the Common European Framework (CEF):

Table 5

Graded Readers

Common European Framework of Reference Levels		IELTS Band scores	General English Examinations	Oxford Bookworms Stages & Headwords	Penguin Readers Stages & Headwords
Upper advanced	C2	7.5 +	CPE		
Advanced	C1	6.5/7.0	CAE	Stage 6 2500	Level 6 3000
Upper intermediate	B2	5.0/5.5/6.0	FCE	Stage 5 1800 Stage 6 2500	Level 5 2300
Intermediate	B1	3.5/4.0/4.5	PET	Stage 3 1000 Stage 4 1400	Level 4 1700
Elementary	A2	3.0	KET& YLE Flyers	Stage 1 400 Stage 2 700	Level 2 600 Level 3 1200
Beginner	A1	-	YLE Movers YLE Starters	Starter 250 Stage 1 400	Easystarts 200 Level 1 300

The rationale for choosing books slightly below their vocabulary size was that when learners knew all, or nearly all, of the words in a text they may have been better able to practise and develop top-down reading skills. This was important for a number of reasons: Learners needed to increase their reading speed in preparation for academic reading; it gave the students the opportunity to experience successful reading; it may have helped to develop a culture of reading where one did not previously exist. Thus, the graded reading was included to address these important aspects of vocabulary and reading development as described in the literature.

A further consideration was that when learners read words they had already practised, but now met in a fuller context, they were given the opportunity to know the words more deeply in terms of use and meaning, as described in table 1 of the Literature Review.

Finally, reading texts composed of already known vocabulary was seen as a way of controlling the variable that the students might increase their vocabulary size through reading as opposed to the retrieval treatment.

Speed Reading

When the students reached a vocabulary size of around the GSL 1st 1,000 words they began Quinn, Nation & Milletts' (2018) speed reading course. The authors state that:

> The programme contains twenty 550 word readings, each with ten comprehension questions. The readings are based on topics related to Asia and the Pacific and are written within the 1000 most frequently used words of English (West, 1953). The only exceptions are words that are explained in the text, the titles of passages or content words like country names and animal names. In addition, the grammar has been restricted by limiting the number of relative clauses, passives and difficult time references. (p.2)

This element was thought to be particularly useful in developing the habit of reading quickly when accessing meaning is not an issue, given that slow reading rates are a major obstacle to success in tests such as the IELTS reading exams.

This activity also aimed to develop the students' ability to decode with a high degree of automaticity. The words in the texts were high frequency and had already been met and retrieved on a number of occasions. As discussed in chapter 2, automaticity in reading appears to be a reliable indicator of performance (Koda, 2005) and may also free up space in working memory, allowing even lower-level readers to devote this to other processes (Crossley, Greenfield & McNamara, 2008).

The course was employed thus:

Class sets of twenty laminated readings and sets of the corresponding questions were prepared for all texts in the course. The aim of developing the skill of reading quickly whilst maintaining understanding was explained to the students. They were asked to read the texts as quickly as possible, knowing that there would be ten comprehension questions to answer afterwards. A record of each student's time, comprehension score and words read per minute (WPM) was kept for every reading text. If a student scored less than eight out of ten on a text any improvement in speed was discounted. A time chart ranging from one to six minutes at ten second intervals was displayed on a smart board together with a timer, after each ten second period expired that time was erased. Once learners had finished reading a text the next time remaining was recorded and the reading text collected. The multiple choice comprehension questions were then made available and answers were recorded. The answer sheets were subsequently marked by peers and checked by the teacher. The twenty text course was run in a period of twenty-five days followed by a gap of four weeks when the course was repeated with the same procedure and in the same sequence.

It should be noted that the course was arranged in a way that was likely to minimise any incidental vocabulary learning; the students had already covered the vocabulary included in the texts repeatedly and the aim of increasing speed would have been severely compromised if a significant amount of vocabulary was unknown at the outset.

Intensive reading

Intensive reading texts were prepared using materials from Macmillan one stop English (2011) which were then adapted to the students' vocabulary size with vocabulary profiler 3.0 which is freely available on Tom Cobb's Lextutor website (2019). The main goal of the intensive

reading was to develop the sub-skills needed to tackle question types such as Reference, Main Ideas, True, False and Not Given, Multiple choice and Comprehension in preparation for the IELTS exam. Little vocabulary work, with the exception of one off pre-teaching of key off-list blocking words, was done by the experimental groups in the context of these reading texts. Again, as the vocabulary level was controlled it was unlikely any new vocabulary would be met in these texts.

Control Groups: Extensive reading

A programme of graded reading was also stipulated for the control groups as seen below for level 3:

The level 3 students were assigned a range of fiction and non-fiction graded readers to read independently, usually at CEFR B1 level or above. After reading they were asked to summarise each book read in their own words. It was hoped this would help develop a habit of reading for pleasure.

The recommended level of graded reader had been set at CEFR B1 for level 3 students which equates to a minimum load of 1,000 headwords and it was suggested that level 2 students should aim for readers of 1,600 headwords by the time of entry to level 3. Given that, in the case of the control groups, the reading load was positioned above the students actual VST it was clear they were, in contrast to the experimental groups, expected to expand their vocabulary through reading.

Control Groups: Intensive reading

The control groups were assigned reading course books for their intensive classroom reading. Level 3 students followed the Reading Explorer 2 (Douglas, 2009) level which is aimed

at elementary to pre-intermediate students. Level 4 students utilised Inside Reading level 1 (Burgmeier & Zimmerman, 2007) supplemented with Achieve IELTS 1 (Harrison & Cushen, 2005). The assumption at level 4 was that students would progress from pre-intermediate level (CEFR A2) to intermediate level (CEFR B1) during the first half of the semester and then move on to the IELTS preparation material in the second half of the semester. Given that Achieve IELTS is designed to take students from strong intermediate (B2) to upper- intermediate (C1) level this appeared to be an optimistic view.

Phase 1: Instruments Used to Measure Vocabulary and Reading Comprehension

The five instruments used to measure vocabulary and reading comprehension were the Vocabulary Size Test (VST); the Vocabulary Levels Test (VLT)**;** the Common Educational Proficiency Assessment (CEPA); the International English Language Testing System (IELTS); and the Preliminary English Test (PET). A table of when each was used is available in appendix F. Other well-known tests were considered for vocabulary testing including the C test, Yes/No tests and the Computer Adaptive Test of Size and Strength (CATSS). The C test has similarities to cloze tests with the main difference in that whereas a cloze test deletes whole words a C test only has the second half of a word deleted. Whilst the C test has a lexical focus which contributes to predicting success in all language skills, it can be seen as a general language proficiency test as opposed to focusing specifically on vocabulary (Grotjahn & Stemmer, 2002). Additionally, C test results seem to correlate with a degree of lexical sophistication which was not thought to be applicable given the level of the students in this study. Yes/No tests (Meara, 2010) were developed to test receptive knowledge of written language. Each level of the tests has sixty items of which 40 are words and sixty non-words. The test taker simply has to mark Y

(I know the word) or N (I do not know the word). The researcher in this study felt it was not clear what 'knowing' a word meant in this context. In view of this and the criticisms of the source lists made by the test developer (Meara, 2010) it seemed that Yes/No tests might measure less accurately than the VLT and VST tests. The CATSS test (Laufer & Goldstein, 2004) was discarded because of its stated aim of testing both breadth and depth of knowledge. This study set out to measure breadth of knowledge of the most frequent English vocabulary with low level students.

Towards the end of this study the researcher became aware of a test in development as the focus of a doctoral study (Kremmel, 2018). This new test seems to hold the promise of better differentiation of vocabulary levels and the ability to accurately predict reading comprehension levels. However, although this test may prove invaluable in future research, it was still in the development process during the current study.

Rationale for the number and spacing of pretest and posttests

The rationale for using multiple posttests was in response to the difficulty in timing tests in order to gauge the effectiveness of an intervention. It was important to chart how quickly any effect took place and also crucial to measure if the effect was lasting and earlier gains consolidated. Equivalent versions of the vocabulary and reading tests were used in order to avoid any effect through familiarity with a specific version (Cohen, Manion & Morrison, 2011).

Delayed posttests in longitudinal vocabulary intervention studies are especially important. The process of learning vocabulary is accumulative; short term studies are, therefore, not ideal. For example, the present study collected VST scores on five occasions consisting of one pretest (before the treatment) and four posttests. The researcher could gauge that the

treatment was having an effect through the use of the word cards which are, in essence, a test. Therefore, the first posttest took place 13 weeks after the pretest; the second eight weeks later including holiday; the third a further five weeks later; and the final posttest seven weeks after the third. This arrangement allowed the researcher to see if the treatment had an effect and, more importantly, if the effect was lasting. The possibility that the posttests could have had some effect on the scores was taken into account. However, given that the treatment itself consisted of continual testing and retrieval, in the form of word cards and online activities, the likely effect of the addition of one or two more tests was considered to be minimal (Schmitt, 2010a)

The instruments which were used are described in the following five sub-sections.

Vocabulary Size Test (VST).

Two versions of the VST were used in this study. The Vocabulary Size Test (VST) was developed by Beglar and Nation (2007) to measure the written receptive vocabulary size of both native users and learners of English. The VST played an important role in the present study allowing initial vocabulary size and growth in vocabulary size to be measured and charted throughout. This test was particularly appropriate in a study measuring growth in written receptive vocabulary growth because it is designed to test knowledge of written word form and knowledge of the relationship between form and meaning.

The sample item below illustrates the format of the VST:

1. SEE: They saw it.
   a. cut
   b. waited for
   c. looked at
   d. started

The form and meaning relationship appears to be an important aspect of the word knowledge needed when reading and, as this study also seeks to discover the strength of the relationship between vocabulary size gain and improvement in reading, the choice of this instrument again appears appropriate. In addition, it tests decontextualised knowledge of a word which corresponds with the type of decontextualised learning the treatment in this study explores (Nation, 2013).

The above is not an argument that vocabulary knowledge alone is sufficient in the quest for reading proficiency. It has been suggested that the role of syntactic knowledge in reading performance has been under investigated especially in terms of research on the comparative importance of syntactic and vocabulary knowledge in reading (Shiotsu & Weir, 2007). Working memory span has been identified as another important factor in reading comprehension. Lee (2014) found that working memory span was a more important factor in reading comprehension than vocabulary for advanced learners although this did not prove true for elementary learners. Whilst these and other factors undoubtedly contribute to reading ability, this study is only concerned with the contribution of vocabulary to reading.

The VST has been shown to be consistent and reliable measure of receptive vocabulary size. Beglar (2010) conducted a Rasch-based validation study of the VST involving 19 native English speakers and 178 Japanese L1 participants. The Japanese participants were divided into 3 proficiency levels: High group (n = 29) (advanced speakers TOEFL scores 560-617); Mid group (n = 53) (intermediate intensive study TOEFL 525; Low group (n = 96) (from lower ranked Japanese university). He concluded that validity was high based on Rasch model

analyses targeting criteria from Messick's (1989; 1995) framework for educational measurement validity. Beglar (2010) found that:

> 1) the items and examinees generally performed as predicted by a priori hypotheses, (2) the overwhelming majority of the items displayed good fit to the Rasch model, (3) the items displayed a high degree of unidimensionality with the Rasch model accounting for 85.6% of the variance, (4) the items showed a strong degree of measurement invariance with disattenuated Pearson correlations for person measures estimated with different sets of items of 0.91 and 0.96, and (5) various combinations of items provided precise measurement for this sample of examinees as indicated by Rasch reliability indices >0.96. (p. 101)

Table 6 indicates that an excellent level of Rasch item reliability (0.96) was demonstrated through the measurement of participants' results on multiple versions of the VST.

The Rasch item separation index was calculated to investigate the possibility of a ceiling effect, when many of the test takers reach or nearly reach the highest score in a test. The results indicate the five forms of the test were not subject to a ceiling effect and were reliable as a measure of increases in vocabulary knowledge. The responsiveness of the VST, whether it can clearly differentiate between clearly different ability levels of test taker, was investigated through the person strata statistic (7.15). This statistic indicates that the VST is able to differentiate written receptive vocabulary knowledge at seven different levels. This ability to track growth in vocabulary size of extended time periods was invaluable in the present study and, as the VST tests vocabulary to the 14,000 or 20,000 word level depending on the version used, there was

little danger of a ceiling effect given that the participants in the present study were elementary level.

Table 6

Item Reliability and Item Separation for Five Forms of the Vocabulary Size Test

Number of items	Number of items per level	Frequency level	Test Takers	Rasch item reliability	Rasch item separation
140	10	1st-14th	ALL	0.96	5.22
80	10	1st-8th	Low & Mid	0.96	4.71
40	5	1st-8th	Low & Mid	0.96	4.93
40	10	1st-4th	Low & Mid	0.98	6.25
20	5	1st-4th	Low & Mid	0.98	6.39

Note. Reprinted from *A Rasch-based validation of the Vocabulary Size Test* by David Beglar, retrieved from https://journals.sagepub.com/doi/10.1177/0265532209340194 Copyright 2010 by Sage Journals

Recent research has questioned whether the vocabulary correctly identified when taking the VST can be accessed by learners when reading (Kremmel & Schmitt, 2016). The authors argue that the VST provides clues to meaning in the form of the options in the multiple choice format above and. as real life reading does not provide clues to meaning, the VST and real reading may not be compatible. They go on to claim that what is tested in the VST is word recognition whilst what is need for reading is automatic recall. However, the present study promoted automatic recall of meaning through the use of word cards. In this context, the participants had no options; they simply saw the target word and recalled the meaning. If this could not be done instantly the target meaning was given and the card replaced in the shuffled

pack until it was later tested again. Participants did not move on to the next set of 20 cards until the current set could all be instantly recalled. Therefore, in this study the VST was confirming what had already been shown when using the word cards. In summary, the VST can be seen as testing the receptive form meaning link identified in the 'types of vocabulary knowledge' (p.26) section of the literature review.

Vocabulary Levels Test (VLT).

Two different versions of the VLT were used in this study. The test was used at the 1,000, 2,000 and 3,000 levels and for the Academic Word list testing. The receptive version of the Vocabulary Levels Test (VLT), (Nation, 1983; Schmitt, Schmitt & Clapham, 2001) was used to provide a further measure of vocabulary size. The VLT was originally designed as a diagnostic test to estimate which of the high frequency English vocabulary students were not familiar with and thus diagnose where their efforts should be concentrated (Nation, 1990; Schmitt, Schmitt & Clapham, 2001). It was thought to be particularly useful in this study because of the anticipated low vocabulary sizes of the participants; the 1000 version of the levels test has contexts which only contain words at the same or higher frequency than the word being tested and, where this proved difficult, pictures were substituted for words (Nation, 1993).

Levels Test (Recognition) Test 2 Instructions: There are 39 questions. Tick (✓)"T" if a sentence is true. Tick (✓)"N" if a sentence is not true. Tick (✓) "X" if you do not understand the sentence.

Example: We can stop time.

T (This is True)

N (This is Not true)

X (I do Not understand the question

Figure 2. VLT 1000 word test format

Source: Nation (1993)

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

l	business	
2	clock	_____ part of a house
3	horse	_____ animal with four legs
4	pencil	_____ something used for writing
5	shoe	
6	wall	

You answer it in the following way.

l	business	
2	clock	___*6*__ part of a house
3	horse	___*3*__ animal with four legs
4	pencil	___*4*__ something used for writing
5	shoe	
6	wall	

Figure 3. Format of other VLTs

Source: Schmitt, Schmitt & Clapham (2001)

Although not designed for the specific purpose of measuring vocabulary size the VLT has often been used in research when this was the aim (e.g., Cameron, 2002; Cobb, 1997; Qian, 2002), especially before the advent of the VST. A reason for not using the test specifically for vocabulary size was that it only tested vocabulary within the specific bands given above (Schmitt, Schmitt & Clapham, 2001). That is it does not test words at the 4,000; 6,000; 7,000; 8,000 and 9,000 frequency levels. However, in this study, given the low vocabulary levels of the participants, it was only used at levels one to three, meaning each level was tested. The VLT is likely to return slightly higher scores than the VST as it is possibly more sensitive to limited but incomplete understanding of a word. This is because the distractors used in the VST share

aspects of meaning with the target word whereas in the VLT, distractors and the correct answer have nothing in common (Beglar & Nation, 2007).

Common Educational Proficiency Assessment (CEPA)

The CEPA was originally designed as a post admission placement test for students entering the foundation programmes of the UAE's federal university sector. This was changed in 2006 to that of a high stakes gatekeeper for entry to the three university sector UAE institutions (Brown & Jaquith, 2011) and was used as such in the present study. The total time limit for the examination, which consists of Grammar and Vocabulary, Reading and Writing papers, is 2 hours. There are no set times for each component although recommendations of 45, 45 and 30 minutes respectively are made (NAPO, 2011). Coombe and Davidson (2014) provide the test specifications as follows

> There are 45 grammar items and 40 vocabulary questions. All the questions on the grammar, vocabulary, and reading sections are multiple-choice. The grammar items measure a candidate's ability to recognize common grammatical patterns in English, and the vocabulary items measure knowledge of common English vocabulary. The Reading section consists of two descriptive or narrative texts of around 400 words in length, and one non-prose text, such as a web page or a brochure, with a total of 25 multiple-choice questions across the three texts. The Writing section consists of an essay task of between 150 and 200 words. The quality of student's writing is assessed in terms of grammar, vocabulary, spelling, and content. Students record their answers and write their essay on an Optical Mark Reader (OMR) sheet and these are later scanned and processed. (p. 270)

There appears to be little independent research on the CEPA and the studies that have been conducted by the national agencies were not made available for this study. Evidence supporting the predictive validity of the CEPA in the context of final first semester foundation results is provided in a doctoral thesis (Rumsey, 2013), however, this Pearson correlation (0.476) was not as strong as previously claimed (0.699) by the National Admissions & Placement Office (NAPO). Rumsey (2013) reports that original NAPO reported correlations are no longer available on their website.

The content validity of CEPA appears to be strong. The university sector end users were closely involved in its development and ensured that their course content informed the CEPA content. Dialogue between the CEPA team and the federal institutions including feedback on any changes to course content continues to inform exam content (Lange, 2012).

In the case of construct validity, Coombe and Davidson (2014) found that the CEPA constructs were sufficiently addressed by the variety of task types used in the examination. The reliability of CEPA is underpinned by the marking system developed by the examination team. All written papers are double blind marked and a third marker is assigned to grade outliers or cases of wide divergence in marks. Data on all markers is collected and used to identify those who are harsh or inconsistent. In addition, pre-exam standardization is carried out online and FACETS software (Linacre, 2019) is used to adjust scores in the light of markers' grading profile. The markers are all qualified to at least Master's level and receive comprehensive training. They are all employed by one of the federal institutions and receive recompense for their efforts (Brown & Jaquith, 2011).

Criticisms of CEPA include the suggestion that it encourages 'teaching to the test' and, in consequence, narrows the range of what is taught. It appears that in many UAE high schools, CEPA has become the curriculum, at least in the final year. Students are required to solely focus on item types likely to occur in the examination to the exclusion of all else. Whilst this may have the effect of improving CEPA grades to some extent, it is unlikely to improve real world English levels. However, these developments are not surprising when one considers that high school teachers, often with inadequate English proficiency, are seen as having sole responsibility for the performance of their students. The likelihood that your employment depends on CEPA results would surely narrow your focus to only exam skills (Rumsey, 2013).

In addition to the above, the authenticity of the CEPA has been questioned based on overuse of multiple choice questions. It was felt that these do not reflect the tasks demanded of undergraduate students although it was conceded that the texts used in the examination have claims to authenticity in the same university context. In their independent review of the CEPA, Coombe and Davidson (2014) called for a research agenda into the CEPA to be constructed. Areas they considered to be key were improving the variety of task types, as raised above, and considering how the testing of listening and speaking might be incorporated.

As of 2016/17 the Emirates Standardized Test (EmSAT) has replaced CEPA as the UAE university entrance test (UAE Ministry of Education, 2018).

International English Language Testing System (IELTS)

The IELTS academic reading test was chosen for this study primarily because it was used as the entrance exam (band 5 needed) for progression to the institution's degree programmes. In other words, the institution stipulated that all students must practice the IELTS tests. The study

used different versions of the academic reading test each time the participants were tested, in order to limit the possibility of the practice effect. The IELTS Handbook (2007), states that the IELTS academic reading test takes a total of sixty minutes. There are 40 questions on three reading texts in the test. The total amount of words in the three texts is between 2,000 and 2,750. The task types used are:

- multiple choice
- short-answer questions
- sentence completion
- note/summary/flow-chart/table completion
- labelling a diagram
- matching headings for identified paragraphs/ sections of the text
- identification of writer's views/claims – yes, no or not given
- identification of information in the text – true, false or not given
- classification
- matching lists/phrases

To interpret any gain in IELTS reading grades in the present study it is first required to define the construct. The IELTS Handbook (2007, p 2) suggests that the Test 'is designed to assess the language ability of candidates who need to study or work where English is the language of communication'. O'Loughlin and Arkoudis (2009) suggest that the academic reading module focuses on general academic literacy. Davies (2008) claims that as the IELTS test is specifically designed as a generic test for those wishing to enter higher education, it purposely avoids any focus on particular academic disciplines in favour of general academic language proficiency. According to Davies (2008) academic language proficiency is:

> …skilled literacy and the ability to move easily across the skills. In other words, it is the literacy of the educated, based on the construct of there being a general language factor relevant to all those entering higher education, whatever specialist subject(s) they will study. (p. 113)

This one size fits all approach has not escaped criticism. Weir, Hawkey, Green, Unaldi and Devi (2009), whilst acknowledging the general academic nature of the IELTS, bemoan the lack of any real expeditious (processing texts quickly and with specific goals) reading tasks. They argue that this is the type of reading university students need to do and if it is not tested how do we know they can? In summary, the IELTS reading module tests academic reading skills but not, perhaps, all of them equally.

The IELTS academic module is used by HCT as a high stakes gatekeeper to assess readiness for entry to undergraduate programmes (HCT, 2012). A plethora of research into the validity, that it tests what it purports to test, of the IELTS academic reading examinations exists (Oliver, Vanderford & Grote, 2012; Coley, 1999; Cotton & Conrow, 1998). Research has shown that the texts used in IELTS academic reading have some differences to those likely to be encountered by undergraduates in terms of cognitive demand. However, many of the features of the IELTS academic texts are comparable to undergraduate texts including vocabulary range and grammatical complexity. Green, Unaldi and Weir (2009), conclude:

> While IELTS passages are at a level of difficulty appropriate to university study, they are not as challenging as some of the texts encountered in the more linguistically demanding areas such as the law textbook analysed for this study. (p. 207)

Notwithstanding the concern noted, IELTS academic reading appears to be a valid test of academic reading.

The reliability of any test, that it produces consistent results, is also extremely important. In this respect the IELTS academic reading module scores highly on internal consistency. The alpha coefficient of the 40-item reading module averaged around 0.90 in extensive analysis of

candidate responses to 16 versions of the 2003 test materials (Blackhurst, 2004); a coefficient of this magnitude is generally regarded as highly reliable (Cohen, Manion & Morrison, 2011: 640).

However, although the academic reading module appears to have strong claims to both reliability and validity, in some respects it may not be the most suitable instrument for measuring the reading improvement of low-level learners. According to the IELTS handbook (2007: 4) a candidate at band one is a 'non-user' and even band three is described as an 'extremely limited user'. The mean entry band level in academic reading of the participants who completed this study (n=74) was 3.1419 with a range of one to five. The academic reading module was designed to measure the performance of candidates wishing to enter English medium higher education and measures most accurately from around band five to band seven. This, together with very low vocabulary size, may have rendered the academic reading module a blunt instrument for these students. As a precaution against this possibility, the Cambridge PET reading test, described below, was also employed in this study.

Preliminary English Test (PET).

This study used different versions of the PET reading test each time the participants were tested, in order to limit the possibility of the practice effect.

The PET examines a variety of reading types in the five parts of the test: 1: careful reading local. Reading real-world notices and other short texts; 2: expeditious reading global; 3: expeditious reading local; 4: careful reading global; 5: careful reading local. The expeditious reading, identified as important by Kalifa and Weir (2009), is mainly restricted to scanning (visual matching) although some search reading (finding information when the specific form of the information is not known) is included. Khalifa and Weir (2009) question whether this

expeditious reading is actually carried out expeditiously as the time spent on each part of the test is not controlled. They suggest that some candidates may always read the text carefully whatever the task.

The Cambridge PET reading module, in common with all Cambridge examinations, has often been examined for reliability and validity (Saville, 2012; Taylor, 2004; Weir, Vidaković & Galaczi, 2013). A particularly detailed account is provided by Khalifa and Weir (2009), who apply a socio-cognitive theoretical framework to the issue of validity and find that Cambridge ESOL have stringent procedures in place to maintain the integrity of their main-suite reading exams.

Cobb (2001; 1995) describes why the PET exam was adopted at a university in Oman. Initially, the university had relied on IELTS testing for evaluating students' readiness to begin degree level studies. The IELTS tests had proven inaccessible with students regularly attaining scores several bands below the levels required to enter more established western universities. PET was chosen as a replacement, as it was hoped the exam might prove less challenging for these students. However, Cobb (1995) reports that the hope was not fulfilled with the majority of students still unable to achieve the newly set bar of the university's own band 4 which corresponded to achieving a PET exam pass grade (passing grades in PET are pass or pass with merit); reading achievement was especially poor. Investigation into possible causes (Cobb, 1995) highlighted the students' low vocabulary sizes; the mean size was around 900 words of the 1st 2000 most frequent after one year of instruction at the university.

This study added the PET reading examination to the testing instruments because it seemed that this examination may be a more sensitive measure of the reading ability of students

with vocabularies of around 2000 words. IELTS academic reading texts appear to be, notwithstanding some differences in word frequency and length, comparable, in terms of vocabulary, to the texts students might encounter in the early stages of an undergraduate degree (Weir et al, 2009). In contrast, the reading texts found in the PET examination are controlled for vocabulary level by means of a word list. Originally, the word list used to this end was the Cambridge English Lexicon (Hindmarsh, 1980) which contained 2,387 words. This list was produced on the foundations of other lists including West's (1953) GSL, Thorndike and Lorge's (1944) list and Kucera and Francis's (1967) list of modern American English. However, by the time of this study the word list for PET had been further developed and had expanded to a total of 2708 words (Street & Ingham, 2007). The development process involved the use of corpora including the British National Corpus and Cambridge's own corpus derived from language candidates had used successfully in main-suite writing exams (Ball, 2002). Despite these changes, the PET lexicon and West's GSL have much in common and, given that the word cards in this study followed West's list, it was felt that PET reading texts were likely to be more accessible for the participants than IELTS academic reading texts. . In summary, the PET should be better suited than IELTS reading for elementary Emirati readers in that it requires efficient word recognition processes but of a smaller vocabulary size.

Phase 1: Statistical Methodology for Experimental Data

The dependent variables measured to address RQ1 and test H1 were the repeated measures for the Vocabulary Size Test (VST), the Vocabulary Level Test (VLT) and the Academic Word List (AWL) scores collected before the intervention (VST & VLT Week 2 in Semester 1; AWL Week 2 in Semester 2) and on multiple occasions after the intervention, until

Week 16 in Semester 2. The repeated measures for the VST scores, termed Test 1 (pretest, before the intervention), Test 2, Test 3, Test 4, and Test 5 (four post-tests after the intervention) were collected on five occasions. The repeated measures for the VLT scores, termed Test 1 (pretest, before the intervention), Test 2, and Test 3 (two post-tests after the intervention) were collected on three occasions. The repeated measures for the AWL scores, termed Test 1, Test 2, were collected on two occasions, specifically Test 1 (pretest) before the intervention (Week 2 in Semester 2) and Test 2 (posttest) 14 weeks after the intervention (Week 16) in Semester 2. The two fixed factors were the group (Control vs. Experimental) and the Level (referring to the two levels of participants classified by their English foundation, divided into low and high, according to their scores for the CEPA English assessment as defined in Table 4.

Multivariate analysis of variance (MANOVA) including repeated measures of multiple dependent variables (i.e., the VST, VLT, and AWL scores) was not appropriate to address RQ1 and test H1 because MANOVA assumed that the number of occasions used to collect the repeated measures of each dependent variable was equal (Huberty & Olenjik, 2006); however, this assumption was violated. Repeated measures Analysis of Variance (ANOVA) with two fixed factors was conducted using the General Linear Model (GLM) procedure in SPSS (Field, 2013) to address RQ1 and test H1. Three tests were conducted, one for the VLT scores, one for the VST scores, and one for the AWL scores. The purpose of repeated measures ANOVA was to determine the effects of the two fixed factors and their interactions on the repeated measures of each dependent variable. Repeated measures ANOVA was applicable, because this method considers the correlations between successive measurements collected over time, by dividing the

data into within-subject and between-subject factors. H1 was supported if the F-test statistics were statistically significant ($p < .05$).

The most important outcome of ANOVA is the ability to estimate and test interaction effects (Hair et al., 2010). When a significant interaction is present, the impact of one factor depends on the value of another factor. The interactions between time x group and time x level were particularly important, because they facilitated the testing of H1 to determine if decontextualised vocabulary study, using word cards and translation, would lead to a more rapid gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element. A significant time x group interaction would imply that the pattern of change in the test scores over time was not the same in the experimental group and the control group. A significant time x level interaction would imply that the pattern of change in the test scores over time was not the same across the two levels of students. A group x level interaction would imply that the differences in the test scores between the groups depended on the levels. Interaction plots to illustrate the relationship between the mean scores vs. time were constructed, with two lines representing the two groups, and two lines representing the two levels. A significant interaction, reflected by one line having a steeper slope than the other line, would determine which students achieved more rapid gains in receptive vocabulary.

Although ANOVA assumes that the dependent variables are normally distributed experiments using simulated data, have revealed that the statistical inferences of ANOVA are robust if the dependent variable deviates from normality (Schmeider, Zeigler, Danay, Beyer, & Buhner, 2001). Although ANOVA assumes equality of variance of the dependent variables across different groups, this assumption is not strongly violated so long as the group sizes are

equal and the GLM procedure with Type III sums of squares is applied (Rutherford, 2001). Furthermore, Levene's test, supported by SPSS to test for equality of variance, may provide unreliable results, and has been described as analogous to sailing a rowing boat on the sea in order to find out if the sea is too rough to sail an ocean liner (McGuinness, 2002). So long as the group sizes are balanced, the sample sizes in each group are sufficiently large to provide adequate statistical power to conduct ANOVA, and Type III sums of squares are applied, then tests for normality and homogeneity of variance are irrelevant (Bowker & Randerson, 2007). Therefore, the results of these tests are not reported.

There is also a considerable amount of objective evidence in the literature to determine the effects of non-normality of the dependent variable on the results of ANOVA and t-tests. This evidence is based on the results of experiments with simulated data (e.g. using the Monte Carlo method). The experimental results have consistently demonstrated that the statistical inferences of parametric statistical tests are robust, even when the data deviate from normality (Glass, Peckham & Sanders, 1972; Harwell, Rubinstein, Hayes & Olds, 1992; Rasch, Kubinger & Moder, 2011; Schmeider, Ziegler, Danay, Beyer & Bühner, 2010).

Furthermore, if the dependent variable is measured at the interval/ratio level (as in the current study), non-parametric tests are not superior to parametric tests when the assumptions of parametric tests are violated (Lix, J. Keselman & H. Keselman, 1996; Rochon, Gondan & Meinhard, 2012; Sawilowsky, 2005; Vickers, 2005; Zimmerman, 2000). When interval/ratio level data are converted into ordinal ranks prior to a non-parametric analysis, the variance in the data is sacrificed, and the results of alternatives to ANOVA (e.g., the Kruskal-Wallis test, or ANOVA by ranks) are underpowered and may be spurious. Despite common belief, non-

parametric tests are not the panacea to solve issues associated with the violation of the assumptions of parametric tests.

In view of the above, the researcher decided than non-parametric testing would not be appropriate for the statistical testing of the experimental data in this study.

The most important assumption of repeated measures ANOVA is sphericity, meaning that the variances of the differences between the repeated measures are equal. If Mauchly's test was significant ($p < .001$) then the assumption of sphericity was violated. If so, them the Huynh-Feldt correction was applied if $\varepsilon > .75$, but the Greenhouse-Geisser correction was applied if $\varepsilon > .75$ (Field, 2013).

The most useful result of ANOVA were not the *p*-values based on the *F* statistics, but the effect sizes, given by partial eta squared (η^2) representing the proportions of the variance in the test scores explained by time, group, and level. The *p*-values provided very unreliable results, because, unlike the effect sizes, the *p*-values were confounded by the sample size, and were meaningless if the sample was not drawn by random sampling from a defined population (Filho, Paranos, da Rocha, Batista, Silva, & Santos, 2013; Halsey, Curran-Everettt, Vowler, & Drummond, 2015; Hubbard & Lindsay, 2008; Nuzzo, 2014; Scucs & Ioannidis, 2017; Wasserstein & Lazar, 2016). The interpretation of the effect size was $\eta^2 < .05$ is negligible; $\eta^2 = .25$ is moderate, and $\eta^2 = .64$ is strong (Ferguson, 2009).

The Bonferroni correction was not applied to reduce the probability of elevated Type I errors when using multiple ANOVA tests (Abdi, 2007). The lowering of the α level to $\alpha = .05/k$, (where k = number of tests) would cause too many Type II errors (Nakagawa, 2004; O'Keefe, 2003; Perneger, 2008).

It was essential to report the effect sizes, to indicate practical significance, as well as the p-values, to indicate statistical significance (American Psychological Association manual, 2010; Hill & Thompson, 2004; Wasserstein & Lazar, 2016). Reporting the effect size, even if the p-value is not statistically significant, is particularly important in experimental research designed to measure the impact of an intervention on a sample. In such experiments, practical significance is what the researcher wants to know (e.g., whether the effect of the intervention on the sample was negligible, moderate or strong). The *p*-value indicates what the researcher does not necessarily want to know (e.g., if the results were caused by sampling error, and if the inferences can be generalized from the sample to a defined population, assuming that the sample was drawn randomly from that population). When the aim of the research is to measure the strength of the effect of an intervention on a non-random sample, then the effect size, not the *p*-value, provides more useful information about what the researcher actually wants to know (Kotrlik & Williams, 2003; Vacha-Haase, 2001).

Phase 1: Modelling Methodology

This section presents the methods of analysis to address RQ2 and test H2 by statistical modelling of the quantitative data. If there was only one measurement of vocabulary size, and one measurement of reading comprehension, then a simple bivariate correlation analysis, using Pearson's correlation coefficient, would be appropriate; however, bivariate correlation analysis was not appropriate because in this study there were repeated measures of receptive vocabulary size obtained from the participants with three vocabulary tests (VST, VLT and AWL) and repeated measures of reading comprehension measured obtained from all the participants with two reading tests (PET and IELTS). Vocabulary size was a latent variable consisting of multiple

VST scores. A latent variable is a multi-faceted construct that cannot be measured with a single score but must be measured with a linear combination of scores that are correlated with each other. Similarly, reading comprehension consisted of linear combinations of multiple PET and IELTS scores. Factor analysis was conducted to determine if the PET and IELTS scores consisted of a single construct (e.g., tapping only one skill), and not two separate constructs (e.g. tapping different skills).

Most of the statistical methods available in SPSS were devised over 50 years ago, and SPSS does not include very modern methods of statistical analysis devised in the last 10 years which can evaluate the correlations between multiple latent variables containing repeated measures (Field, 2013). Partial least squares structural equation modelling (PLS-SEM) using SmartPLS software was therefore applied to address the second research question. The reasons for using PLS-SEM were: (a) PLS-SEM is a very modern method that permitted the repeated measures of the test scores to be linearly combined into latent variables using composite factor analysis, which is best suited to answering RQ2 of this study; (b) PLS-SEM is a non-parametric method that is not sensitive to the distributional and measurement characteristics of the variables; and (c) the results could be visualised using a path diagram. When using PLS-SEM, the correlations between two or more latent variables can be evaluated using a single model visualised in the form of a path diagram (Hair, Hult, Ringle, & Sarstedt, 2017). The analysis was conducted using SmartPLS software, as described by Wong (2013). A CSV (comma-delimited) file containing the data was imported into SmartPLS. A path diagram was drawn using the graphic user interface of SmartPLS, as depicted in Figure 4.

The oval blue symbols are the two latent variables, specifically Vocabulary Size and Reading Comprehension. The multiple rectangular yellow symbols are the indicators, specifically the test scores which were linearly combined to measure each latent variable. Each indicator was labelled according to its source (i.e., VST, VLT, IELTS, or PET) and its time of measurement (i.e., 1 to 5). The arrows pointing out of the latent variables into the indicators represented the loading coefficients (λ) computed by composite factor analysis.

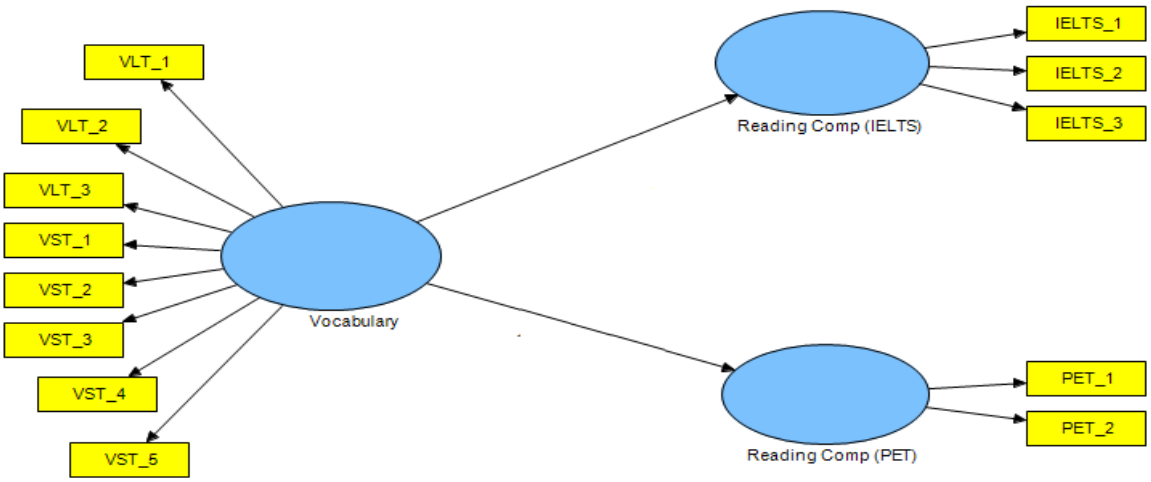

Figure 4. Path diagram drawn using SmartPLS

Composite factor analysis is a modern 21st century method of factoring that uses completely different statistics and algorithms to validate factors than other classical factoring methods, such as Principal Components Analysis (PCA) which were developed over 50 years ago (Hair et al., 2017; Rigdon, 2012). Composite factor analysis did not extract information from the covariance matrix, nor did it involve any form of rotation, and unlike PCA, it assumed that

the latent variables were correlated (Afthanhorhan, 2013). The λ coefficients, which could potentially range from 0 to 1, were interpreted to evaluate the validity of the latent variables. If all the loading coefficients were strong ($\lambda > 0.5$) then the factorial validity of the latent variables was confirmed. If, however, the loading coefficients were weak (< 0.5) then the latent variables were not valid, and the model was untenable (Hair et al., 2017). This interpretation of λ was not consistent with alternative methods of factor analysis (e.g., using principal components) where $\lambda > 0.7$ is generally used to indicate factorial validity (Hair et al., 2010).

The arrow between the latent variables represented the path coefficients (β) which could range from -1 through 0 to + 1. The β coefficients indicated the relative strength and direction (positive or negative) of the correlation between vocabulary size and reading comprehension. The β coefficients were interpreted to test hypothesis H2: The receptive vocabulary size of the elementary Emirati learners correlate with the PET reading scores. In contrast, the IELTS reading scores correlate only with the receptive vocabulary size of those participants who exhibited the greatest receptive vocabulary gains. The effect size was given by the R^2 statistic, representing the proportions of variance in Reading Comprehension explained by Vocabulary Size. The interpretation of the R^2 statistic was < .05 = negligible; .25 = moderate; .64 = strong (Ferguson, 2009). If $R^2 > .25$ the relationships between Vocabulary Size and Reading Comprehension exhibited a high-level of practical significance in the context of educational research (Fraenkel & Wallen, 2011).

Limitations of Phase 1

It is extremely important to discuss in considerable detail the limitations of the quantitative part of this study. This discussion is imperative, and cannot be abbreviated, because quantitative studies in education, involving the use of descriptive and inferential statistical analysis, based on data collected in surveys and experiments, have been very severely criticized in the literature for over 25 years (Carver, 1993; Daniel, 1998; Lohman, 2006; Maxwell, 2004; May, Boe & Boruch, 2003; Menon, 1993; Mittag & Thompson, 2000; Nix & Barnette, 1998; Schanzenbach, 2012; Smeyers & Depaepe, 2010; Yilmaz, 2013). Even though millions of dollars have been spent by the educational research establishment on quantitative intervention studies in the last 25 years, no new effective innovative practices have been identified, and no new evidence-based educational policies have been implemented, that have significantly improved the reading abilities of the neediest of students (Pogrow, 2017). Pogrow criticises the quantitative methods used by many researchers to assess the effectiveness of educational practices to improve reading ability as having flawed and exaggerate actual effectiveness. Experimental studies do not provide the type of information practitioners need, therefore research on effective practices tends to mislead rather than inform practice. House (2008, p. 637) similarly criticized the use of quantitative experimental studies, specifically randomized controlled trials, to evaluate the impact of educational interventions because 'they have produced little of substance'. Furthermore, Serduyuk (2017) complained that, even after 50 years of quantitative research to evaluate the impact of educational innovations:

> We realize that the actual pace of educational innovations and their implementation is too slow, as shown by the learning outcomes of both

> school and college graduates, which are far from what is needed in today's world. (p.5)

It is therefore essential to comprehensively review and fully understand the limitations of Phase I of this study, in order to justify the choice of the explanatory mixed methods research design as a more effective approach than conducting only quantitative research to evaluate the effectiveness of an innovative educational intervention.

The first limitation of the quantitative phase of this study was the sample size. If the sample size was too small, then Type II errors may occur (i.e., the differences between the mean scores based on the sample data may be declared to be not statistically significant, whereas in the population, these differences may, in fact, be statistically significant). A power analysis to estimate the minimum total sample a Repeated Measures ANOVA was conducted using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) with the following input data: Effect size = low (0.15); medium (0.25) and large (0.5); statistical significance level (α) = .05; power ($1 - \beta$) = 0.8; number of groups = 2; number of repeated measures = 5; moderate correlation among repeated measures = 0.5; moderate non-sphericity correction = 0.5. The minimum total sample size, assuming an equal number in each group was N = 98 participants for a low effect size; N = 34 participants for a medium effect size; and N = 10 participants for a large effect size. Therefore, the total sample size of N = 40 used in this study, with equal group sizes of n = 20 in the control group and n = 20 in the experimental group, was adequate to conduct Repeated Measures ANOVA assuming a medium to large effect size.

A power analysis was also conducted to estimate the minimum total sample size to conduct a t-test using the quantitative survey data with the following input data: Effect size = low

(0.15); medium (0.25) and large (0.5); statistical significance level (α) = .05; power ($1 - \beta$) = 0.8. The minimum total sample size was 277 participants for a low effect size; 102 participants for a medium effect size; and 27 participants for a large effect size. The total sample size of N = 106 participants in the survey and N = 40 participants in the experiment were adequate to conduct *t*-tests assuming a medium to large effect size.

A power analysis was not necessary for the analysis of the data with PLS-SEM, because this method achieves very high power at low sample sizes, due to bootstrapping. The bootstrap involved the computation of mean values by randomly sampling and re-sampling the data with replacement using 5000 sub-samples drawn from the data, based on the Monte Carlo algorithm. Hair et al. (2010) asserted that PLS-SEM 'is insensitive to sample size considerations. Its estimation approach handles both very small and very large sample sizes…particularly useful for generating estimations even with very small sample sizes (as low as 30 observations or less)' (p.776).

Separate to the above sample size considerations, the experimental research design using repeated measures of the examination scores used in this study had weaknesses because it was underpinned by assumptions that might be violated. The threats to validity included selection bias, the maturation effect, the Hawthorne effect, the diffusion effect, the ecological fallacy, and the regression fallacy (Fraenkel & Wallen, 2011).

Selection bias was a potential threat to validity because intrinsic differences between the abilities and other characteristics of the students assigned to the experimental and control groups before the experiment was conducted could potentially influence the outcomes (Fraenkel & Wallen, 2000). Consequently, independent samples t-tests were conducted with SPSS to

determine if the mean VST, VLT, and AWL scores were significantly different between the control and experimental groups at Time 1.

Maturation would be a threat to internal validity if the participants were young children whose cognitive skills, knowledge and attitudes change naturally as they grow up over a period of several years, and these changes are not associated with a prescribed educational intervention (Fraenkel & Wallen, 2011). Because the time period of this study was short (less than one year), and the participants were adults, the maturation threat was considered to be negligible.

The Hawthorne Effect, also known as reactivity, occurs when being asked to take part in an experimental intervention alerts or sensitises the participants in such a way that they alter their natural behaviour (Fraenkel & Wallen, 2011). Therefore, the repeated measures of the test scores of some of the experimental group may be biased because the participants reacted to the knowledge that they were part of an experiment. The Hawthorne effect also refers to some of the control group members feeling that they must work harder than normal to ensure that the expected superiority of the experimental group is not demonstrated. The Hawthorne effect is not easy to measure or control in practice, because it depends upon the personal attitudes of the participants, that may differ from one sample to another. The Hawthorne effect was probably not a source of bias in this study, because the students were not asked to confirm whether they wanted to be members of the experimental group or the control group. The participants were not informed in advance of the detailed objectives or the expected results, and they were not made to feel obliged to react in a positive way to the effects of the intervention. Therefore, the Hawthorne effect or reactivity was probably not a limitation of the results of this study, because the students did not actually realise that they were participating in an experiment.

Diffusion (i.e., the transfer of information by social interaction between one group and another group) was also probably not a limitation because it was unlikely that the control and experimental groups came together socially to discuss the extent to which decontextualised vocabulary study, using word cards and translation, could lead to greater gain in receptive vocabulary for elementary level Arab learners of English than a similar teaching program lacking this element.

The results and conclusions of this study, and all other studies in educational settings involving the statistical analysis of multiple test scores collected over time could potentially be limited by the regression effect. This effect is caused by random measurement error, and natural fluctuations that take place over time, irrespective of the effect any type of prescribed intervention or educational program. The regression effect results in lower scores at the first test tending to become higher and move up toward the overall mean score at the final test, whereas higher scores at the first test tend to become lower and move down toward the overall mean score in subsequent tests (Nesselroade, Stigler, & Bates, 1980). The outcome of the regression effect is that the results of Repeated Measures ANOVA might not reflect the direct impact of the decontextualised vocabulary study, using word cards and translation, leading to an overall gain in receptive vocabulary for elementary level Arab learners of English. The results of ANOVA could be caused indirectly by the regression effect. An improvement in the test performance of students may be assumed by an educational researcher to be directly caused by a prescribed intervention or educational program. However, in fact, this apparent improvement may be caused indirectly by the regression effect. This error is called the regression fallacy (Weeks, 2007). The regression fallacy is a common source of error in educational research and has resulted in the

misinterpretation of the results of tests to determine the impact of educational interventions (Lohman, 2006).

The quantitative results of this study could also be potentially limited by the ecological fallacy, meaning that each individual member of a group does not behave in exactly the same way as the mean value computed for the whole group. Therefore, statistical inferences based on mean test scores deduced from the aggregated test performance data collected from a group of individuals do not provide any information whatsoever about the test performance of each individual who is a member of that group. The ecological fallacy is common source of error in educational research, when students are often classified into groups, and mean values are used to summarise and compare the performance of each group (May, Boe, & Boruch, 2003).

The limitations discussed above do not imply that the quantitative methodology implemented by the researcher was fundamentally incorrect, or that the results of the quantitative study were meaningless. Many similar limitations are common to most previous research to evaluate the effects of interventions in educational settings that involve the use and interpretation of inferential statistics (Fraenkel & Wallen, 2011; Schanzenbach, 2012; Smeyers & Depaepe, 2010).

Yilmaz (2013) reviewed arguments suggesting that qualitative research (i.e., any type of research in educational settings that generates findings not arrived at by statistical procedures or other means of quantification) may often provide superior results to quantitative research. Therefore, the results of qualitative research (e.g., based on thematic analysis of interview transcripts to elucidate the perceptions of students) may be more helpful than quantitative research to raise awareness of the potential factors that may influence the effectiveness of a

prescribed intervention (Maxwell, 2004). Applying the constructivist paradigm, learning performance should not only be assessed quantitatively using the test scores of students assigned by the researcher into groups. The assessment of learning performance should also focus on the subjective performance of each unique individual student, taking into account qualitative contextual factors, such as his or her culture, literacy, language, personal interests, needs, and social interactions (Hein, 2007; Liu & Matthews, 2005; Mogashoa, 2014).

The use of a sequential explanatory mixed methods research design, involving the enrichment and elaboration of quantitative data using qualitative data helped to overcome the limitations of conducting only quantitative research in educational settings (Fraenkel & Wallen, 2011). The following sections describe how the application of mixed methods helped the researcher to achieve 'constructive realism' (Cupchick, 2001) and 'pragmatism' (Creswell, 2014) to address the research questions.

Phase 2: Qualitative Methodology

The results of Phase 2 were interpreted mainly to address RQ3: What are the perceptions of elementary level Emirati learners of English regarding the learning of vocabulary and its relationship to reading comprehension? The results of Phase 2 were also interpreted to enrich the answers to RQ1: To what extent, and in what ways, does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? and RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores?

The researcher interviewed a total of 21 Emirati learners of English drawn from the target population. To protect their rights to anonymity and confidentiality each student was identified with an alphanumeric code, which defined his level (3 or 4) and group (C = Control, or E = Experimental). It should be noted that the interviews took place at the end of the second semester of the study, in the second semester foundation level two of semester one became level three and foundation level three became level four. These students are the original participants who started the study; followed by a student ID number (S1 to S6) as summarised in Table 7.

Table 7

Code Names of Interview Participants

Group	Level	Code Names of Participants					
Control	3	3 C S1	3 C S2	3 C S3	3 C S4	3 C S5	
Experimental	3	3 E S1	3 E S2	3 E S3	3 E S4	3 E S5	
Control	4	4 C S1	4 C S2	4 C S3	4 C S4	4 C S5	
Experimental	4	4 E S1	4 E S2	4 E S3	4 E S4	4 E S5	4E S6

The researcher conducted semi-structured interviews, meaning that the interviews were only loosely based on a series of predefined open-ended questions. For example: ‘How do you think the exam was?’; ‘Do you think one exam was more difficult that the other?’; ‘How did you understand that question?’; ‘What did you do in the exams when you saw the words and you didn't know them?’; ‘Why is the Wordpower Dictionary, the online one, why is that good?’; ‘How much do you read when you are not at college?’. However, for much of the time, the researcher intervened with prompts in order to encourage the Emirati learners of English to elaborate upon their answers and describe in richer detail their own experiences of learning vocabulary and the problems they experienced in reading.

Computer assisted qualitative data analysis software (CAQDAS) such as Nvivo, could potentially be used by the researcher to interpret the interview transcripts. Thematic analysis refers to the extraction of themes, where a theme consists of a phrase, sentence, or group of sentences that identify a single unit of meaning (Merriam, 2014). Although the researcher attempted to use Nvivo to extract themes, he found that many of the themes did not appear to be meaningful. The literature also confirms that CAQDAS involves too many mechanistic and rigid processes and puts pressure on the researcher to focus on volume and breadth, rather than on depth of meaning (St. John & Johnson, 2000). The use of CAQDAS tends to make the researcher exceed the limits of the valid conclusions that can be drawn from qualitative data, for example, through the coding of an excessive number of themes that may have limited meaning (Shönfelder, 2011). NVivo and all other CAQDAS are only data management packages. Although CAQDAS facilitates the search for and extraction of themes the researcher must always remain in control. CAQDAS cannot properly analyse qualitative data, because this is the primary task of the researcher. CAQDAS does not confirm the scientific value of qualitative research. It simply makes the handling of qualitative data slightly easier (Zamawe, 2015).

In order to extract in-depth meanings from the interview transcripts that could not be obtained by use of CAQDAS, the researcher needed to make sense of the meaning of the qualitative data by personally exploring, manually extracting, and interpreting the significant statements of the participants to identify emergent themes that would help to support the triangulation process. Instead of using CAQDAS software, all of the significant statements provided by each participant, which could be classified into themes, were transcribed line by line into the rows of an MS Excel spreadsheet. The framework used to extract themes from the

interview transcripts was similar to that recommended by Maguire and Delahunt (2017) for researchers in teaching and learning. This framework involved the six steps listed in Table 8, based on an original framework designed by Braun & Clarke (2006). Table 8 did not define a simple linear series of steps. The researcher had to move backward and forward through the first four steps before the primary and secondary themes could ultimately be defined and the results could be tabulated.

Table 8

Six Stage Framework for Thematic Analysis

Step 1: Become familiar with the data
Step 2: Generate initial codes
Step 3: Search for themes
Step 4: Review and refine themes
Step 5: Define primary and secondary themes
Step 6: Tabulate the results

Braun & Clark's (2006) framework above distinguishes between two types of themes, termed semantic and latent. Semantic themes are extracted from the 'bottom up', meaning that they do not look for anything beyond interpreting what the participants said, and the themes are not underpinned by any other predefined sources of information, such as a literature review. Latent themes, in contrast, are extracted from the 'top down', meaning that they look beyond what was said and are based on a predefined template based on concepts and theories derived from a literature review. Semantic themes were considered by the researcher to be more relevant than latent themes, because the purpose of the thematic analysis was to enrich the quantitative

data, and to perform a triangulation, but not to expand existing concepts and theories in the literature.

The interview transcripts contained a total of 395 statements, including the interventions of the researcher, which were not included in the thematic analysis. Table 9 shows that a total of 153 significant statements were extracted from the responses of the students to identify five primary themes.

Table 9

Primary Themes from Interview Transcripts

Primary theme	Frequency	Coverage (% of total number of themes)
1: Reading	55	36.2
2: Vocabulary	44	28.9
3: Exam	37	24.3
4: Speaking and Listening	9	5.9
5: Writing	7	4.6

Each significant statement consisted of verbatim phrases or sentences spoken by a student to identify a named unit of meaning that could be interpreted to address the research questions. All irrelevant statements were excluded. Each student's significant statement was coded with the student's level (3 or 4); the student's group (C = Control or E = Experimental); and the student's code name (S1 to S6). The sort function of MS Excel was used to aggregate the coded statements into primary themes, as described by Meyer & Avery (2009). The primary themes containing the highest frequencies of statements, and the highest coverage (%), were Theme 1: Reading (55, 36.2%); Theme 2: Vocabulary (44, 28.9%); and Theme 3: Exam (37, 24.33%). Theme 4:

Speaking and Listening (9, 5.2%) and Theme 5: Writing (7, 4.0%) had the least coverage. Each primary theme was subsequently refined into sub-themes, each of which expressed a slightly different manifestation of the primary theme.

Phase 3: Triangulation

Figure 1 highlights that the final component of the sequential explanatory design included the triangulation of the qualitative and quantitative data. Triangulation in educational research, otherwise known as cross-examination or cross-verification, has been applied for over twenty years to help improve the credibility of the findings of mixed methods research through the collective interpretation of qualitative and quantitative data (Fraenkel & Wallen, 2011; Denzin, 1997). Guion (2002) classified triangulation into five categories (a) Data triangulation, which involves comparing and contrasting data across time, space, and/or persons; (b) Investigator triangulation, which uses multiple rather than single observers to record the same event; (c) Theory triangulation, which employs different theories to explain the findings; (d) Environmental triangulation, which uses different locations and settings related to the environment in which the study took place; and (e) Methodological triangulation, which seeks to find convergence and corroboration of findings obtained using different methods. Adopting Guion's classification, the current study applied Methodological triangulation, because both quantitative and qualitative methods of data collection and analysis were applied to address common research questions, and the results obtained using both methods were compared.

With specific regard to research in education, triangulation, as applied in the current study, usually refers to the application and combination of two or more methods in the study of the same phenomenon in an educational setting (Fraenkel & Wallen, 2011; Johnson &

Christensen, 2004). Methodological triangulation in educational research is posited to overcome the weaknesses and biases which may arise from the use of only one method. If the findings obtained using qualitative and qualitative methods converge, and lead to the drawing of similar conclusions, then the conclusions are more likely to be credible (Denzin, 1997).

Integration is a more modern and complex approach for the collective interpretation of quantitative and qualitative sources of data, particularly when the challenge is to provide deeper insights into complex, multilevel processes and systems (Creamer, 2018; Moran-Elis, Alexander, & Cronin, 2006). Integration is therefore most applicable for mixed methods research in medicine/healthcare (Cathain & Murphy, 2010; Fetters, Curry, & Creswell, 2013) and in management/organisational research (Tunarosa & Glynn, 2016). Integration refers to an overall assessment of the goodness of fit of the quantitative findings to the qualitative findings and generally includes three processes: (a) confirmation; (b) expansion; and (c) discordance. Confirmation, which is equivalent to triangulation in educational research (as described above), seeks to determine if the findings of qualitative research converge with, or corroborate, the findings of quantitative research. Expansion seeks to determine how the findings of quantitative and qualitative research help to increase knowledge by addressing a single phenomenon from different epistemological perspectives. Discordance seeks to determine the extent to which the quantitative findings are inconsistent, incongruous, contradict, conflict, or disagree with the qualitative findings. The researcher adopted the discordance element of integration but found that the qualitative and quantitative findings were in agreement. The expansion component of integration was not applicable to the current mixed methods study because the researcher aimed

only to use confirmation or triangulation in order to determine the extent to which the findings of Phase 2 converged with or corroborated the findings of Phase 1.

Limitations of Phase 2

The relatively small sample size of *N* = 21 students who attended the interviews was not necessarily a serious limitation for Phase 2. Unlike quantitative methodologies, qualitative methodologies do not require a large representative sample of participants to provide data that can be generalised to a population, or a sample that is large enough to provide the power to achieve statistical significance (Bogdan & Biklen, 2006; Creswell, 2014; Merriam, 2014). The sample size of N = 21 student used in Phase 2 of this study was chosen to achieve saturation, meaning that more is not necessarily better when collecting qualitative data by interviews. A larger sample size does not necessarily provide more themes than a smaller sample size. For example, Green and Thorogood (2009) stated that 'the experience of most qualitative researchers is that, in interview studies, little that is new comes out of transcripts after you have interviewed 20 or so people' (p. 120). Mason (2010) examined the sample sizes in a total of 516 studies that used qualitative interviews for data collection and suggested that about 15 participants was the minimum sample size to achieve saturation. Therefore, the sample size of N = 21 participants used in this study was acceptable for the purpose of thematic analysis.

The main limitation of Phase 2 was the reliability of the interpretation of the interview responses which were conducted in English. It is possible that the thematic analysis could potentially be contaminated by researcher bias, meaning that the researcher may have distorted the results, in order to portray a desired outcome. Consequently, the researcher needed to apply reflexivity, meaning that in order to 'negotiate the swamp' of qualitative data analysis (Finlay,

2002, p. 209) and avoid researcher bias, he had to reflect personally upon what role he played in constructing knowledge, and tried to understand how and why he interpreted the qualitative data in such a way as to extract certain themes in preference to others, and apply these themes to enrich the quantitative data (Day, 2012; Johns, 2004; Johnson & Duberley, 2003). Because the researcher was closely involved in the teaching and interviewing of the participants, he may have unconsciously biased the results of the thematic analysis towards providing favourable answers to the research questions (e.g., to conclude that decontextualised vocabulary study, using word cards and translation, resulted in greater gain in receptive vocabulary). In order to determine the extent to which the results of the thematic analysis were contaminated by researcher bias, a reliability analysis was conducted.

Cohen's Kappa (K) was computed to estimate the reliability of the thematic analysis of the interview data. The justification for using K was that, despite its critics (Gwet, 2002), K is widely used in medical, social science, and educational research to evaluate the level of agreement between two raters (Agresti, 2013). In order to estimate K the two independent raters must classify a series of items (e.g., statements provided by interview participants) into several categories (e.g. themes). The K statistic is measured on a scale ranging from +1 (perfect agreement) through 0 (no agreement above that expected by chance) down to -1 (complete disagreement). The value of K was interpreted as follows: $K \leq 0$ indicated no agreement; K = .01 to .20 indicated none to slight agreement, K = .20 to .40 indicated fair agreement, K = 41 to .60 indicated moderate agreement, K = .61 to .80 indicated substantial agreement and K = .81 to 1.00 indicated almost perfect to perfect agreement.

The reliability analysis was conducted by two independent raters, specifically the researcher (Rater 1) and another teacher and researcher of English as a foreign language (Rater 2). The two raters independently extracted sub-themes from a total of 59 statements provided by the students, all of which were classified within the primary theme "Vocabulary" by Rater 1. The cross-tabulation used to compute Cohen's K using SPSS is presented in Table 10.

Nine themes, coded from 1 to 9, were identified by Rater 1, whilst Rater 2 identified six themes, coded from 1 to 6. Rater 2 failed to identify three themes (coded by 0) that were coded 7, 8, and 9 by Rater 1. The proportion of statements that both raters agreed should be included within the same theme (defined by "Yes" in Table 10) was 45/59 = 76.3%. Despite the difference, the results of the reliability analysis reflected substantial agreement between the two raters ($K = .729$, $p < .001$). Consequently, the researcher suggests that the thematic analysis of the interview statements was reliable.

Table 10

Cross Tabulation of Themes Used in Reliability Analysis

Statement	Primary Theme	Rater 1 Sub-theme (Text code)	Rater 2 Sub-theme (Text code)	Rater 1 Sub-theme (Numerical code)	Rater 2 Sub-theme (Numerical code)	Agreement (Yes or No)
32	Vocabulary	Dictionary	Dictionary	2	2	Yes
34	Vocabulary	Dictionary	Dictionary	2	2	Yes
36	Vocabulary	Dictionary	Dictionary	2	2	Yes
38	Vocabulary	Dictionary	Dictionary	2	2	Yes
40	Vocabulary	Teacher	Teacher	5	5	Yes
44	Vocabulary	Dictionary	Dictionary	2	2	Yes
46	Vocabulary	Dictionary	Dictionary	2	2	Yes
52	Vocabulary	Dictionary	Dictionary	2	2	Yes
89	Vocabulary	Difficult	Difficulties	1	1	Yes
92	Vocabulary	Teacher	Teacher	5	5	Yes
95	Vocabulary	Words to Know	Words to Know	4	4	Yes

97	Vocabulary	Translation	Not identified	9	0	No
99	Vocabulary	Dictionary	Dictionary	2	2	Yes
101	Vocabulary	Dictionary	Dictionary	2	2	Yes
103	Vocabulary	Dictionary	Dictionary	2	2	Yes
125	Vocabulary	Teacher	Teacher	5	5	Yes
130	Vocabulary	Number of words	Not identified	8	0	No
131	Vocabulary	Number of words	Not identified	8	0	No
142	Vocabulary	Difficult	Difficulties	1	1	Yes
144	Vocabulary	New words	Not identified	7	0	No
154	Vocabulary	Difficult	Difficulties	1	1	Yes
158	Vocabulary	New words	Not identified	7	0	No
159	Vocabulary	Difficult	Difficulties	1	1	Yes
160	Vocabulary	Difficult	Difficulties	1	1	Yes
166	Vocabulary	Cards	Word Cards	6	6	Yes
170	Vocabulary	Cards	Word Cards	6	6	Yes
171	Vocabulary	Cards	Word Cards	6	6	Yes
173	Vocabulary	Cards	Word Cards	6	6	Yes
183	Vocabulary	Cards	Word Cards	6	6	Yes
185	Vocabulary	Words to Know	Words to Know	4	4	Yes
187	Vocabulary	Words to Know	Words to Know	4	4	Yes
195	Vocabulary	Words to Know	Words to Know	4	4	Yes
197	Vocabulary	Cards	Cards	6	6	Yes
204	Vocabulary	Difficult	Difficulties	1	1	Yes
206	Vocabulary	New words	Not identified	7	0	No
210	Vocabulary	New words	Not identified	7	0	No
211	Vocabulary	New words	Not identified	7	0	No
213	Vocabulary	New words	Not identified	7	0	No
214	Vocabulary	New words	Not identified	7	0	No
217	Vocabulary	New words	Not identified	7	0	No
248	Vocabulary	Difficult	Difficulties	1	1	Yes
306	Vocabulary	Cards	Word Cards	6	6	Yes
307	Vocabulary	Cards	Cards	6	6	Yes
315	Vocabulary	Speed reading	Speed reading	3	3	Yes
316	Vocabulary	Speed reading	Speed reading	3	3	Yes
318	Vocabulary	Speed reading	Speed reading	3	3	Yes
320	Vocabulary	Speed reading	Speed reading	3	3	Yes
321	Vocabulary	Speed reading	Speed reading	3	3	Yes
323	Vocabulary	Dictionary	Dictionary	2	2	Yes
324	Vocabulary	Dictionary	Dictionary	2	2	Yes
326	Vocabulary	Dictionary	Dictionary	2	2	Yes

327	Vocabulary	Dictionary	Dictionary	2	2	Yes
331	Vocabulary	New words	Not identified	7	0	No
332	Vocabulary	New words	Not identified	7	0	No
334	Vocabulary	New words	Not identified	7	0	No
336	Vocabulary	Difficult	Difficulties	1	1	Yes
389	Vocabulary	Speed reading	Speed reading	3	3	Yes
393	Vocabulary	Cards	Word Cards	6	6	Yes
395	Vocabulary	Dictionary	Dictionary	2	2	Yes

Rater 2 was asked by the researcher why he did not identify all of the themes that were identified by the researcher. Rater 2 replied that 'When conducting interviews in qualitative research, you must only ask open ended questions, and you must not ask closed ended questions that can only be answered by 'Yes' or 'No'. Furthermore, you must not put words into the mouths of the participants. For example, you asked the question 'How many words do you think you need to know to read English well? OK. I would say most of you, know well, somewhere between 2000 and 3500 words'. All of the students in level 4 of the control group answered 'Yes'. I believe that your question attempted to put words into the students' mouths, and their answers are not reliable, but reflected acquiescent response bias. This means that the participants probably agreed with you and said 'Yes' only because they deferred to your authority, rather than provide truthful answers'.

The researcher agreed with Rater 2, that acquiescent response bias may have influenced the reliability of the thematic analysis. Furthermore, the literature indicates that acquiescent response bias may be a consequence of forcing respondents to choose certain responses (Ferrando, 2010); and may also be an aspect of the cultural communication style of Arab respondents (Smith, 2004). Consequently, the researcher did not identify any other themes in the student's responses that were based on the answers to closed-ended questions, or answers

that could be answered only with ‘Yes’ or answers that could potentially be biased by acquiescent responding. This meant that after the inter-coder check, Rater 2 's coding was made to be the final coding. The researcher reviewed the data once again to amend the coding based on the recommendations of Rater 2.

Limitations of Phase 3

The triangulation implemented in Phase 3 of the sequential explanatory design could be limited because it might not automatically determine the extent to which the findings of Phase 2 were consistent with, converged with, corroborated, or confirmed the findings of Phase 1. The limitation of triangulation is that quantitative and qualitative methods access different types of data based on different ontological and etymological perspectives. For example, quantitative data analysis assumes that facts and feelings are separate, whereas qualitative data analysis assumes that facts and feelings are not separate. The juxtaposition of the interpretation of facts vs. feelings, and the opposing elements of positivism vs. constructivism, may invite contradiction and tension rather than consistency, convergence, corroboration, and confirmation (Creswell & Plano-Clark, 2011; Moran-Elis, et al., 2006; Morgan, 2014).

Ethical Considerations

The privacy and confidentiality of participants were maintained throughout the research process in line with the university’s ethics policy. All new foundation students took part in an orientation week, when the nature of the new approaches to foundation teaching was explained. The potential participants were met on day one of the semester and the research aims and structure of the study were explained. They were made aware that data collected would be used

to compare two approaches to vocabulary learning and to investigate to what extent there was a relationship between vocabulary size gain and reading skills development. It was further explained that taking part in the study would not require any extra time commitment and that whether they took part or not would not affect their grades or relationship with the HCT. They were then provided with consent and information sheets in English and Arabic (see Appendix D) and asked to read these carefully before making their decision on participation. Time for questions to the researcher was made available and a bi-lingual member of faculty was present in case of comprehension issues.

Signed consent forms were received from all participants (see Appendix D). The students were assured that their identities would not be made public and that reference in the study would be made to their number not name. Additionally, they were also assured that any data gathered or personal information would only be used to meet the research aims, and were informed of their right to withdraw from the study at any time without needing a reason. All participants were assigned a group and student number to ensure individual identities remained secure.

CHAPTER 4: Results

Introduction

The results of this mixed methods study are presented systematically in four sections: (1) Phase 1: Survey; (2) Phase 1: Experiment; (3) Phase 1: Modelling; (4) Phase 2 abbreviations for the statistics followed the guidelines in the American Psychological Association (2010) manual.

Phase 1: Survey

The mean and standard deviation were computed to summarise the 5-point responses to each of the questionnaire items listed in Appendix A and Appendix B. Independent samples t-tests were also conducted to determine if the mean scores for 13 of the items were significantly different between the control group and the experimental group at the conventional $\alpha = .05$ level. The effect size (Cohen's *d*) was also computed to provide an estimate of the practical significance of the difference between the two mean scores. Cohen's d was interpreted using the criteria defined by Cohen (1992) as follows: 0.20 = small effect; 0.50 = moderate effect; and 0.80 = strong effect.

The Bonferroni correction was not applied to reduce the probability of Type I errors (a Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the significance level, and is often denoted by α.) when using multiple tests (Abdi, 2007) where α is conventionally reduced to .05/number of tests. The lowering of the α level below .05 creates Type II errors (a Type II error occurs when the researcher accepts a null hypothesis that is false. The probability of committing a Type II error is called Beta, and is often denoted by β. The probability of not committing a Type II error is called

the Power of the test and provides unreliable results (Nakagawa, 2004; O'Keefe, 2003; Perneger, 2008). The reason for not correcting for Type I errors was that real-world application of the outcomes of a hypothesis test (and not statistical theory or convention) determines whether the researcher is more accepting of Type I or Type II errors. The most important thing that a researcher must do, when conducting research that has practical applications in educational or medical science, is to minimize the false negative (Type II) error rate (Szucs & Ioannijdis, 2017). In this study, the researcher had an obligation to protect the students who participated in this study from Type II errors, regarding the evaluation of (a) the extent to which decontextualised vocabulary study, using word cards and translation, contributed to the learning of receptive vocabulary; and (b) the extent to which the receptive vocabulary size was more strongly correlated with the PET scores than the IELT scores. Type II errors would mean that findings obtained by the researcher that could ultimately lead to beneficial social change (e.g. improvements in the education of elementary Emirati learners of English) are not accepted by the scientific community. Researchers who report results that are not statistically significant due to Type II errors (even though the results might have practical significance) are often blamed for conducting a poorly designed experiment (Maxwell & Kelley, 2011). Results that are not statistically significant are the cause of so called 'publication bias' or the 'file drawer' problem, meaning that the results are believed to be useless, and are therefore hidden and filed away without being made available to other scientists (Dickersin, 1990; Fanelli, 2012; Rosenthal, 1992; Scargle, 2000). Because academic journals tend to publish only statistically significant results (generally only when $p < .05$) the published evidence tends to overstate the impact of interventions, even if the effect sizes are very small (Simonshon, Nelson, & Simmons, 2014).

Therefore, the practical significance of the results (i.e., effect sizes) must always be reported, irrespective of whether the p-values indicate that tests are statistically significant or not statistically significant (Hill & Thompson, 2004; Kotrlik & Williams, 2003; Vacha-Haase, 2001).

Tables 11 and 12 present summaries of the survey data, based on a descriptive and inferential statistical analysis of the 5-point scale responses to the two questionnaires (see Appendix A and B). The responses of the control group vs. the experimental group were compared. The effect sizes (Cohen's *d)* are included for all the *t*-test results, complying with the recommendation that effect sizes should always be provided, even for results of tests in educational research that are not statistically significant ((Hill & Thompson, 2004)

Table 11

Descriptive Statistics for Survey Questions about Learning Vocabulary

Learning vocabulary by:	Control group (n = 71)		Experimental group (n = 35)		*M* difference	*t* (104)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*				
Reading	3.14	1.13	2.97	0.89	0.17	0.78	.438	0.17
Exercises	3.42	1.01	3.74	1.07	-0.32	-1.51	.135	0.31
Communication	4.11	1.02	4.00	0.97	0.11	0.54	.588	0.11
Online learning	3.55	1.09	4.34	0.84	-0.79	-3.78	<.001*	0.82
Vocabulary books	3.42	1.17	3.37	1.17	0.05	0.21	.832	0.04
Vocabulary lists	3.45	1.24	3.51	0.92	-0.06	-0.27	.789	0.06
English only dictionary	3.38	1.07	3.23	1.11	0.15	0.68	.507	0.14
English/Arabic dictionary	4.03	0.99	4.09	0.92	-0.06	-0.29	.773	0.06
Word Cards			4.31	0.93				

Table 12

Descriptive Statistics for Survey Questions about Difficulties with Reading

Learning vocabulary by:	Control group (n = 71)		Experimental group (n = 35)		M difference	t (104)	p	Cohen's d
	M	SD	M	SD				
Reading speed	3.55	1.04	3.63	1.06	-0.08	-0.37	0.714	0.08
Unknown vocabulary	4.03	1.06	4.06	1.03	-0.03	-0.13	0.894	0.03
Text layout	2.13	1.28	1.66	0.97	0.47	1.92	0.057	0.42
Topics	3.37	0.99	3.06	1.08	0.31	1.47	0.146	0.30
Grammar	2.97	1.08	2.99	0.94	-0.02	-0.13	0.896	0.03

The responses to the survey questions indicated that the members of the experimental group generally agreed that learning vocabulary using Word Cards was very useful (M = 4.31, SD = 0.93). The responses to the survey questions also indicated that reading was perceived to be the least useful for learning vocabulary by both the control group (M = 3.14, SD = 1.13) and the experimental group (M = 2.97, SD = 0.89). The control group perceived that communication (M = 4.11, SD = 1.02) and English/Arab dictionary (M = 4.03, SD = 0.99) were the most useful ways to learn vocabulary. The experimental group perceived that online learning (M = 4.34, SD = 0.84); word cards (M = 4.31, SD = 0.93); and English/Arab dictionary (M = 4.09, SD = 0.92) were the more useful for learning vocabulary. The independent samples t-tests indicated that there were no significant differences ($p > .05$) between the mean scores for seven different methods of learning vocabulary with respect to the two groups of students. There was, however, a significant difference between the mean scores for online learning (M = -0.79). The mean score for the control group (M = 3.55, SD = 1.09) was significantly less (t (104) = -3.78, $p < .001$) than the mean score for the experimental group (M = 4.34, SD = 0.84). Applying the criteria defined

by Ferguson (2009) the effect size (Cohen's d = 0.82) reflected that the mean difference between the control and experimental groups for online learning had practical significance as well as statistical significance. The effect sizes for the other seven methods of learning vocabulary (Cohen's d = 0.04 to 0.31) reflected that the mean differences had limited practical significance.

The mean scores < 3.0 to the survey questions about reading difficulties indicated that both the control group (M = 2.13, SD = 1.28) and the experimental group (M = 1.66, SD = 0.97) generally disagreed that text layout (i.e., reading English from left to right) was a source of reading difficulty. The mean scores < 3.0 also indicated that both the control group (M = 2.97, SD = 1.08) and the experimental group (M = 2.99, SD = 0.94) generally disagreed that grammar was a source of reading difficulty. The mean scores > 4.0 indicated that the control group generally agreed that they had most reading difficulties due to unknown vocabulary (M = 4.03, SD = 1.06). The experimental group also generally agreed that they had most reading difficulties due to unknown vocabulary (M = 4.06, SD = 1.03). The independent samples t-tests indicated that there were no significant differences (p > .05) between the mean scores for the five sources of reading difficulty with respect to the control and experimental groups. The effect sizes (Cohen's d = 0.03 to 0.42) reflected that these differences had limited practical significance.

Phase 1: Experiment - Descriptive Statistics

This section presents the statistical evidence to address the first research question: To what extent and in what ways, does decontextualised vocabulary study, using word cards and translation, lead to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? The descriptive statistics for the VST, VLT, and AWL scores used to measure vocabulary size are summarised in Table 13,

classified by the two groups (experimental vs. control). The total number of students was $N = 74$ in each test, with $n = 37$ in the experimental group and $n = 37$ in the control group, apart from the two AWL tests (AWL-1 and AWL-2), that were attended by $n = 18$ high-level students in each group and no low-level students.

Table 13

Descriptive Statistics for VLT, VST, and AWL Scores by Group

Test	Experimental Group			Control Group		
	n	*M*	*SD*	*n*	*M*	*SD*
VLT-1	37	68.43	13.02	37	72.43	12.76
VLT-2	37	80.92	13.03	37	76.73	10.24
VLT-3	37	84.96	7.96	37	78.35	8.60
VST-1	37	4.32	1.86	37	5.65	1.58
VST-2	37	6.57	1.77	37	6.16	1.76
VST-3	37	7.22	1.73	37	5.81	1.37
VST-4	37	8.19	1.87	37	7.76	1.42
VST-5	37	12.68	2.63	37	10.08	2.42
AWL-1	18	18.56	7.79	18	18.22	5.95
AWL-2	18	45.94	11.66	18	39.44	10.75

The mean VLT scores of the students in the experimental group increased over three repeated measures (VLT-1 to VLT-3) from 68.43 to 84.96, representing a mean gain of 16.51. During the same time period, the mean VLT scores of the control group increased from 72.43 to 78.35, representing a smaller mean gain of 5.92. The mean VST scores of the students in the experimental group increased during the five repeated measures (VST-1 to VST-5) from 4.32 to 12.68, representing a mean gain of 8.35. During the same time period, the mean VST scores of the students in the control group increased from 5.65 to 10.08, representing a smaller mean gain of 4.43. The mean AWL scores of students in the experimental group increased over two

repeated measures (AWL-1 and AWL-2) from 18.56 to 45.94, representing a mean gain of 27.39. During the same time period, the mean AWL scores of the level 4 students in the control group increased from 18.22 to 39.44, representing a smaller mean gain of 21.22.

The descriptive statistics for the VST, VLT, and AWL scores classified by the high and low levels are summarised in Table 14.

Table 14

Descriptive Statistics for VLT, VST, and AWL Scores by level

Test	Low-level			High-level		
	n	*M*	*SD*	*n*	*M*	*SD*
VLT-1	38	65.00	14.16	36	76.17	8.51
VLT-2	38	72.76	11.44	36	85.22	8.44
VLT-3	38	78.29	9.44	36	85.19	6.68
VST-1	38	4.00	1.69	36	6.03	1.36
VST-2	38	5.61	1.44	36	7.17	1.73
VST-3	38	5.68	1.51	36	7.39	1.46
VST-4	38	7.13	1.28	36	8.86	1.57
VST-5	38	10.45	2.75	36	12.36	2.60
AWL-1				36	18.39	6.83
AWL-2				36	42.69	11.53

The total number of students was $N = 74$ in each test, with $n = 38$ classified as low-level and $n = 36$ classified as high-level, apart from the last VST test (VST-6) and the two AWL tests (AWL-1 and AWL-2), that were attended by $n = 36$ high-level students, and no low-level students. The mean VLT scores of the low-level 3 students increased over three repeated measures (VLT-1 to VLT-3) from 65.00 to 78.29, representing a mean gain of 13.29. During the same time period, the mean VLT scores of the high-level 4 students increased from 76.17 to 85.19, representing a mean gain of 9.02. The mean VST scores of low-level students increased

during the five repeated measures (VST-1 to VST-5) from 4.00 to 10.45, representing a mean gain of 6.45. During the same time period, the mean VST scores of the high-level students increased from 6.03 to 12.36, representing a mean gain of 6.33. The mean AWL scores of the high-level students over two repeated measures (AWL-1 and AWL-2) increased from 18.39 to 42.69, representing a mean gain of 24.31.

Phase 1: ANOVA - VLT Scores

The VLT scores were analysed to address RQ1: To what extent and in what ways, does decontextualised vocabulary study, using word cards and translation, lead to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? An independent samples *t*-test indicated that the mean VLT score of the experimental group at Time 1 (M = 68.43, SD = 13.02) and the mean VLT score of the control group at Time 1 (M = 72.43, SD = 12.76) were not significantly different (t = 1.33, p = .186). Therefore, the vocabulary level of the two groups was equivalent before the intervention. A repeated measure ANOVA was conducted to determine the effects of the group and the level on the VLT scores. Mauchly's test was significant (W = .626, $p < .001$) so the assumption of sphericity was violated. The Huynh-Feldt correction was applied because $\varepsilon > .75$. The ANOVA tables are presented in Table 15 (within-subject effects); and Table 16 (between-subject effects).

The within-subject effects were (a) the VLT scores increased significantly with time (F = 52.70, $p < .001$) with a moderate effect size ($\eta^2 = 0.43$) and (b) there was a significant time x group interaction (F = 12.01, $p < .001$) with a small effect size ($\eta^2 = 0.15$). The interactions

between time x level and time x level x group were not significant, indicated by $p > .05$ for the F statistics with negligible effect sizes.

The interaction plot in Figure 5 reflects the time x group interaction. The effect of time was not the same for both groups. In the experimental group, the mean VLT score increased rapidly between Time 1 and Time 2 and also between Time 2 and Time 3. In the control group the mean VLT score increased more slowly between Time 1 and Time 2, and also between Time 2 and Time 3.

Table 15

Within-subject Effects for VLT Scores

Effect		Type III SS	df	MS	*F*	*p*	η^2
Time	Sphericity Assumed	4998.52	2.00	2499.26	52.70	<.001*	0.43
	Greenhouse-Geisser	4998.52	1.46	3434.53	52.70	<.001*	0.43
	Huynh-Feldt	**4998.52**	**1.54**	**3241.53**	**52.70**	**<.001***	**0.43**
	Lower-bound	4998.52	1.00	4998.52	52.70	<.001*	0.43
Time x Group	Sphericity Assumed	1138.88	2.00	569.44	12.01	<.001*	0.15
	Greenhouse-Geisser	1138.88	1.46	782.54	12.01	<.001*	0.15
	Huynh-Feldt	**1138.88**	**1.54**	**738.56**	**12.01**	**<.001***	**0.15**
	Lower-bound	1138.88	1.00	1138.88	12.01	.001*	0.15
Time x Level	Sphericity Assumed	312.30	2.00	156.15	3.29	.040	0.05
	Greenhouse-Geisser	312.30	1.46	214.59	3.29	.056	0.05
	Huynh-Feldt	**312.30**	**1.54**	**202.53**	**3.29**	**.053**	**0.05**
	Lower-bound	312.30	1.00	312.30	3.29	.074	0.05
Time x Group x Level	Sphericity Assumed	120.88	2.00	60.44	1.27	.283	0.02
	Greenhouse-Geisser	120.88	1.46	83.06	1.27	.276	0.02
	Huynh-Feldt	**120.88**	**1.54**	**78.39**	**1.27**	**.278**	**0.02**
	Lower-bound	120.88	1.00	120.88	1.27	.263	0.02

Note: * Significant ($p < .05$)

Table 16

Between-subject Effects for VLT Scores

Effect	Type III SS	df	MS	F	p	η^2
Group	296.55	1.00	296.55	1.51	.224	0.02
Level	5743.90	1.00	5743.90	29.19	.001*	0.29
Group x Level	198.39	1.00	198.39	1.01	.319	0.01

Note: * Significant ($p < .05$)

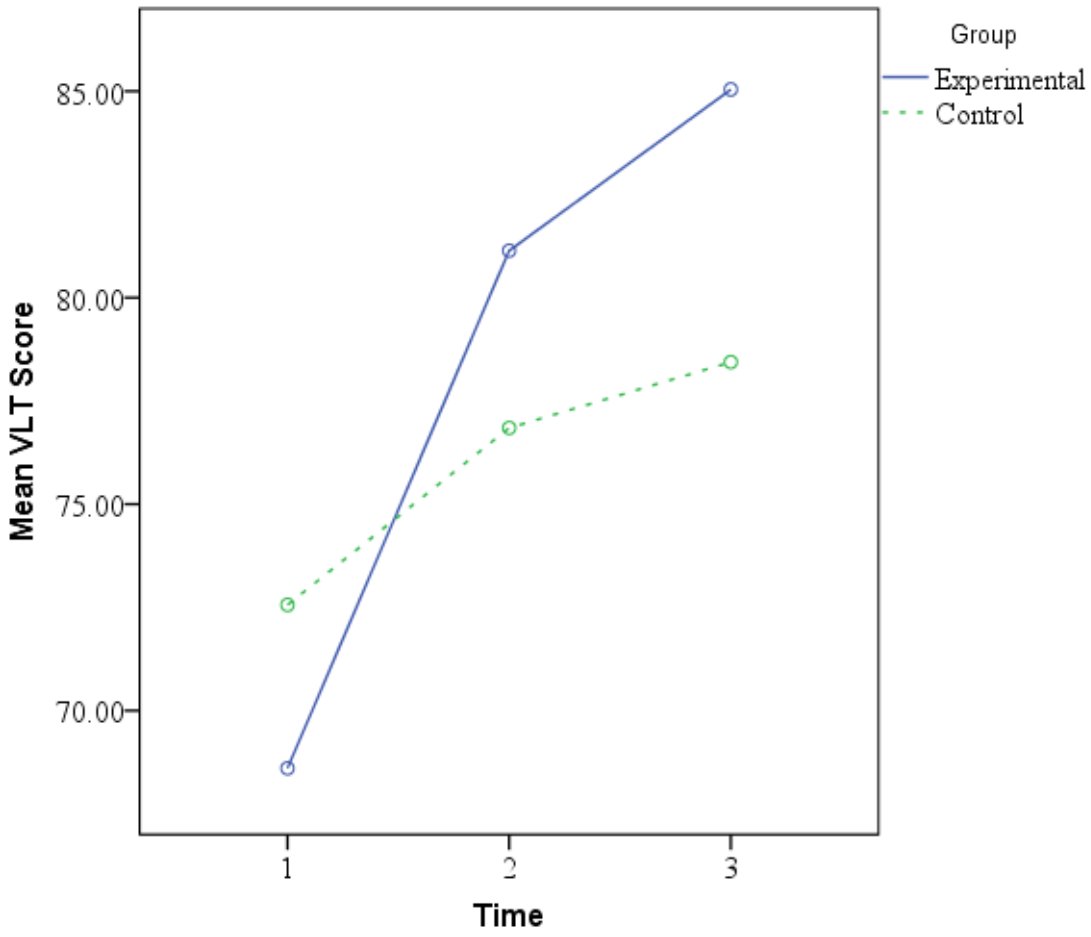

Figure 5. Time x group interaction in mean VLT scores

The between-subject effect indicated that the VLT scores were significantly different with respect to level ($F = 29.19$, $p < .001$) with a moderate effect size ($\eta^2 = .29$). The group had

no significant effect, and the interaction between time x group, was also not significant (indicated by $p > .05$ for the F statistics, with negligible effect sizes).

Phase 1: ANOVA - VST Scores

The VST scores were analysed to address RQ1: To what extent and in what ways, does decontextualised vocabulary study, using word cards and translation, lead to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? An independent samples t-test indicated that the mean VST score of the experimental group at Time 1 ($M = 4.32$, $SD = 1.86$) and the mean VST score of the control group at Time 1 ($M = 5.65$, $SD = 1.58$) were significantly different ($t = 3.34$, $p < .001$). Therefore, the vocabulary size of the two groups was not equivalent before the intervention. This violated the assumptions of repeated measures ANOVA, which assumes that the scores of the participants are equivalent at the baseline A repeated measures ANOVA was conducted to determine the effects of the group and the level on the VST scores, in order to address RQ1. Mauchly's test was significant ($W = .390$, $p < .001$) so the assumption of sphericity was violated. The Greenhouse-Geisser correction was applied because $\varepsilon < .75$. The ANOVA tables are presented in Table 17 (within-subject effects); and Table 18 (between-subject effects).

The within-subject effects were (a) the VST scores increased significantly with time ($F = 247.06$, $p < .001$) with a strong effect size ($\eta^2 = 0.78$) and (b) there was a significant time x group interaction ($F = 21.42$, $p < .001$) with a moderate effect size ($\eta^2 = 0.23$). The interactions between time x level, and time x level x group were not significant ($p > .05$ for the F statistics and negligible effect sizes). The interaction plot in Figure 6 reflects the time x group interaction. In the experimental group, the VST scores increased more rapidly with time than the control

group. In the experimental group, the mean scores increased most rapidly between Time 1 and Time 2, and between Time 4 and Time 5. In the control group, the mean scores tended to increase slowly between Time 1 and Time 2 but most rapidly between Time 3 and Time 5.

Table 17

Within-subject Effects for VST Scores

Effect		*Type III SS*	*df*	*MS*	*F*	*p*	η^2
Time	Sphericity Assumed	1762.44	4.00	440.61	247.06	<.001*	0.78
	Greenhouse-Geisser	**1762.44**	**2.68**	**658.06**	**247.06**	**<.001***	**0.78**
	Huynh-Feldt	1762.44	2.91	604.77	247.06	<.001*	0.78
	Lower-bound	1762.44	1.00	1762.44	247.06	<.001*	0.78
Time x Group	Sphericity Assumed	152.83	4.00	38.21	21.42	<.001*	0.23
	Greenhouse-Geisser	**152.83**	**2.68**	**57.06**	**21.42**	**<.001***	**0.23**
	Huynh-Feldt	152.83	2.91	52.44	21.42	<.001*	0.23
	Lower-bound	152.83	1.00	152.83	21.42	<.001*	0.23
Time x Level	Sphericity Assumed	2.50	4.00	0.62	0.35	.844	0.01
	Greenhouse-Geisser	**2.50**	**2.68**	**0.93**	**0.35**	**.766**	**0.01**
	Huynh-Feldt	2.50	2.91	0.86	0.35	.783	0.01
	Lower-bound	2.50	1.00	2.50	0.35	.556	0.01
Time x Group x Level	Sphericity Assumed	6.87	4.00	1.72	0.96	.428	0.01
	Greenhouse-Geisser	**6.87**	**2.68**	**2.57**	**0.96**	**.404**	**0.01**
	Huynh-Feldt	6.87	2.91	2.36	0.96	.409	0.01
	Lower-bound	6.87	1.00	6.87	0.96	.330	0.01

Note: * Significant ($p < .05$)

Table 18

Between-subject Effects for VST Scores

Effect	*Type III SS*	*df*	*MS*	*F*	*p*	η^2
Group	45.18	1.00	45.18	6.78	.011*	0.09
Level	295.31	1.00	295.31	44.33	<.001*	0.39
Group x Level	1.62	1.00	1.62	0.24	.623	0.00

Note: * Significant ($p < .05$)

The between-subject effect was that (a) the VST scores were significantly different with respect to the group ($F = 6.78$, $p = .011$) but with a small effect size ($\eta^2 = .09$) and (b) the VST scores were significantly different with respect to the level ($F = 44.33$, $p < .001$) with a moderate effect size ($\eta^2 = .39$). There was no significant group x level interaction (indicated by $p > .05$ for the F statistic, and a negligible effect size).

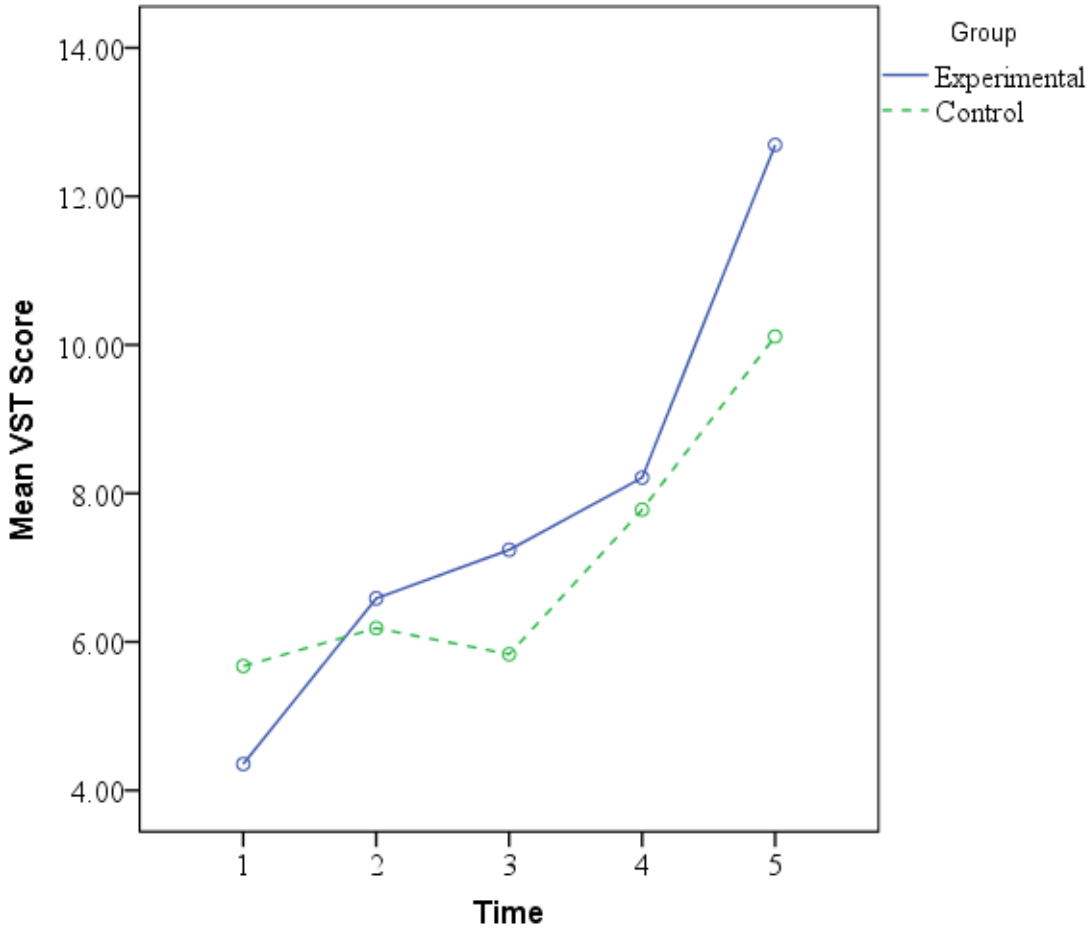

Figure 6. Time x group interaction in mean VST scores

Phase 1: ANOVA - AWL Scores

The AWL scores were analysed to address RQ1: To what extent and in what ways, does decontextualised vocabulary study, using word cards and translation, lead to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? An independent samples *t*-test indicated that the mean AWL score of the experimental group at Time 1 (M = 18.56, SD = 7.79) and the mean AWL score of the control group at Time 1 (M = 18.22, SD = 5.95) were not significantly different (t = 0.15, p = .804). Therefore, the academic word level of the two groups was equivalent before the intervention. A repeated measures ANOVA was conducted to determine the effects of the group on the AWL scores. Level was not included in this analysis because the AWL scores were obtained only for the high-level 4 students. Mauchly's test was not relevant because there were only two repeated measures. The ANOVA tables are presented in Table 19 (within-subject effects); and Table 20 (between-subject effects).

Table 19

Within-subject Effects for AWL Scores

Effect	Type III Sum of Squares	df	Mean Square	*F*	*p*	η^2
Time	10633.68	1.00	10633.68	404.55	<.001*	0.92
Time x Group	171.13	1.00	171.13	6.51	.015*	0.16

Note: * Significant (p < .05)

Table 20

Between-subject Effects for AWL Scores

Effect	Type III Sum of Squares	df	Mean Square	*F*	*p*	η^2
Group	210.13	1.00	210.13	1.43	.241	0.04

The within-subject effects were (a) the AWL scores increased significantly with time ($F = 404.55$, $p < .001$) with a strong effect size ($\eta^2 = 0.92$) and (b) there was a significant time x group interaction ($F = 6.51$, $p = .015$) with a small effect size ($\eta^2 = 0.16$). The interaction plot in Figure 7 reflects this interaction.

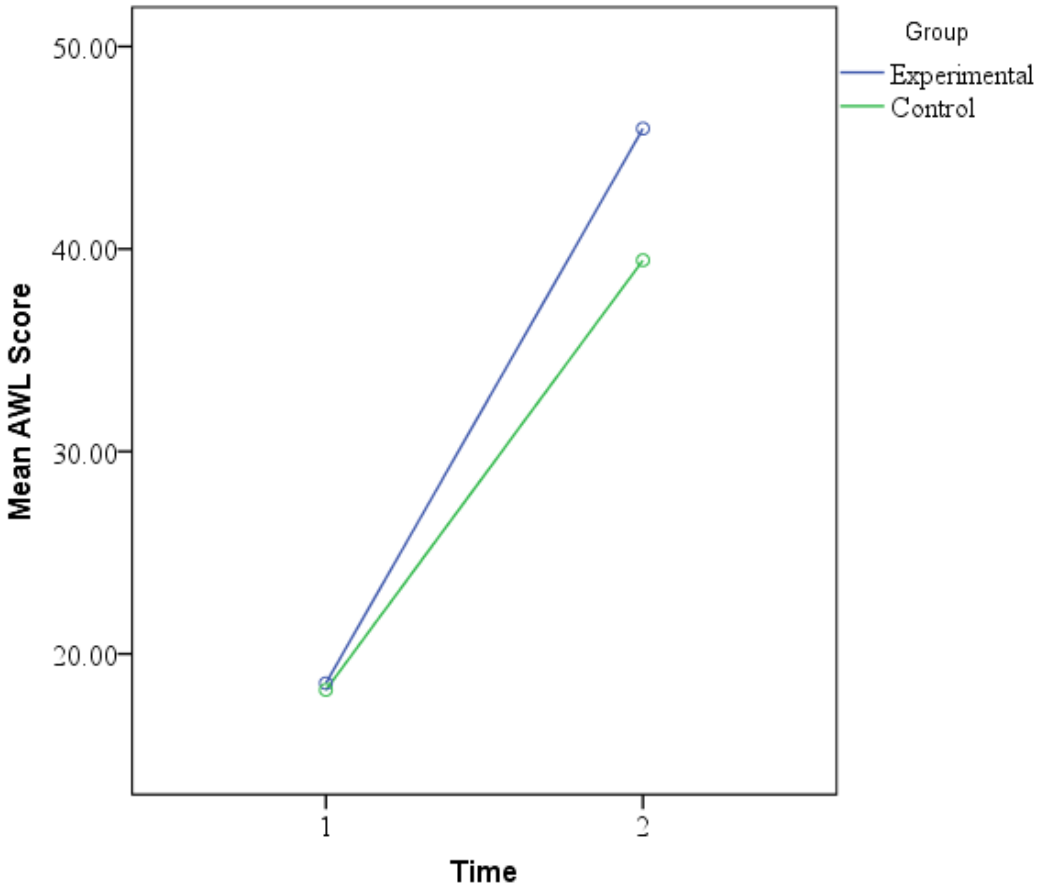

Figure 7. Time x group interaction in mean AWL scores

The two lines in Figure 7 were not parallel, indicating that the AWL scores in the experimental group increased more rapidly with time than in the control group. The between-subject effect was that the AWL scores were not significantly different with respect to the group ($F = 1.43$, $p = .241$) with a negligible effect size ($\eta^2 = 0.04$).

In conclusion, the experiment conducted in Phase 1 provide sufficient statistical evidence to partially support the stated hypothesis, H1: Decontextualised vocabulary study, using word cards and translation, may lead to a more rapid gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element.

Phase 1: Modelling

This section presents the statistical evidence to address the RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores?

Table 21 presents the descriptive statistics of the IELTS reading test scores for the 37 students in the experimental group and the 37 students in the control group. Table 22 presents the descriptive statistics of the PET reading test scores for the 37 students in the experimental group and the 37 students in the control group. In the experimental group, the mean IELTS scores increased from 3.12 to 3.96, whereas in the control group the IELTS scores increased from 3.16 to 3.92. The mean PET scores increased from 16.14 to 21.19 in the experimental group, whereas in the control group the mean PET scores increased from 16.22 to 18.84.

Table 21

Descriptive Statistics for IELTS Reading Test Scores

Test	Group					
	Experimental			Control		
	M	*n*	*SD*	*M*	*n*	*SD*
IELTS_1	3.12	37	0.78	3.16	37	0.75
IELTS_2	3.28	37	0.51	3.32	37	0.50
IELTS_3	3.96	37	0.57	3.92	37	0.48

Table 22

Descriptive Statistics for PET Reading Test Scores

Test	Group					
	Experimental			Control		
	M	*n*	*SD*	*M*	*n*	*SD*
PET_1	16.14	37	4.87	16.22	37	3.74
PET_2	21.19	37	5.32	18.84	37	4.43

To determine if the IELTS, PET, VLT, and VST scores measured the same or different constructs, a composite factor analysis was conducted using SmartPLS. Composite factor analysis did not extract information from the covariance matrix, nor did it involve any form of rotation (Afthanhorhan, 2013). Consequently, no statistics to indicate the goodness of fit of the model to the covariance matrix, and no information about the method of rotation are provided.

Table 23 presents the composite factor loading coefficients, which measured the correlations between the item scores and the factors. The composite factor analysis revealed three factors. The factor loading coefficients (λ) and their cross-loadings for the three factors representing latent variables (Factor 1 = Vocabulary measured with VLT and VST scores; Factor 2 = Reading Comprehension with IELTS scores; and Factor 3 = Reading Comprehension measured with PET scores) are compared.

Table 23

Factor Loading Coefficients for Vocabulary and Reading Scores (IELTS and PET)

	Factor 1 Vocabulary	Factor 2 Academic Reading (IELTS)	Factor 3 General Reading (PET)
IELTS_1	.615	**.766**	.404
IELTS_2	.291	**.625**	.360
IELTS_3	.537	**.810**	.466
PET_1	.532	.519	**.876**
PET_2	.814	.504	**.949**
VLT_1	**.727**	.540	.545
VLT_2	**.791**	.560	.579
VLT_3	**.748**	.469	.520
VST_1	**.615**	.766	.404
VST_2	**.759**	.542	.517
VST_3	**.825**	.427	.653
VST_4	**.792**	.478	.648
VST_5	**.651**	.242	.665

Note: Factor loadings in **bold** indicate the factors to which the IELTS, PET, VLT, and VST items were most strongly correlated.

All of the loading coefficients used to identify each factor were strong ($\lambda > 0.5$) complying with the quality criteria for composite factor analysis, which are lower than alternative methods of factor analysis, such as $\lambda > 0.7$ for principal components factor analysis (Hair et al., 2017). The VST and VLT scores were most strongly correlated with Factor 1: Vocabulary ($\lambda = .615$ to .825). The IELTS scores were most strongly correlated with Factor 2: Reading Comprehension (IELTS) ($\lambda = .625$ to .810). The two PET scores were most strongly correlated with Factor 3: Reading Comprehension (PET) ($\lambda = .876$ to .949). The consistently strong loading coefficients (> 0.7) reflected factorial validity.

The lower cross-loadings of the items for alternative factors confirmed that Vocabulary, Academic Reading (IELTS) and General Reading (PET) were mutually exclusive constructs. The results of the composite factor analysis confirmed the discriminant validity of the three latent variables and indicated that Reading Comprehension consisted of two constructs, probably because they tapped into different skills. The composite reliability coefficients were .906 for Vocabulary, .780 for IELTS and .909 for PET, reflecting that the three latent variables exhibited good internal consistency.

Figure 8 presents the results of the PLS-SEM output by the graphic user interface of SmartPLS, based on the analysis of the VLT, VST, IELTS, and PET scores of $N = 74$ participants (including both the control and experimental groups).

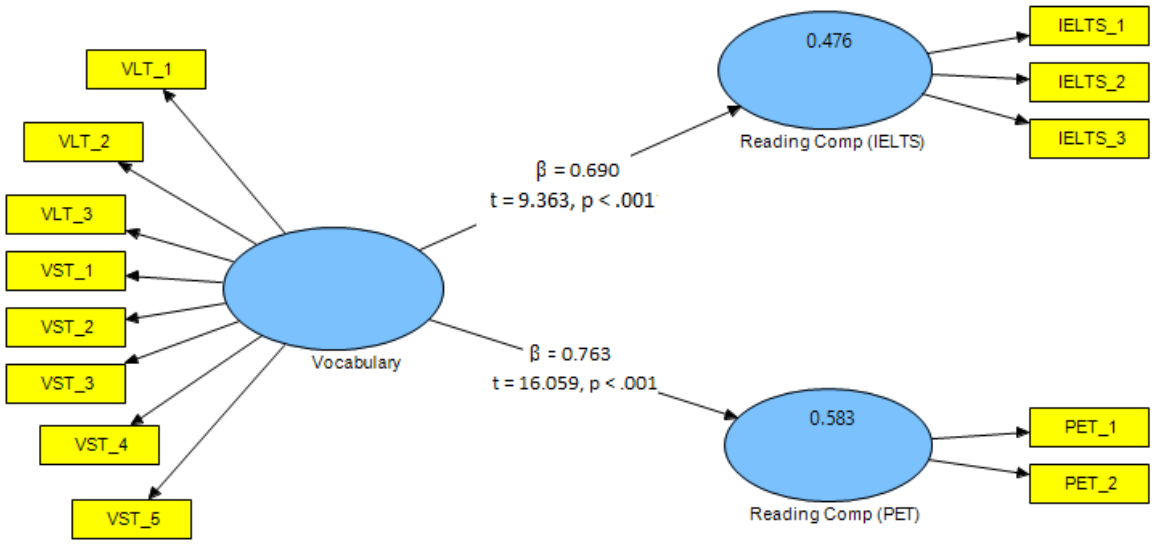

Figure 8. PLS-SEM path diagram with path coefficients

Note: The vocabulary size was more strongly correlated with the PET scores (β = .763) than the IELTS scores (β = .690)

The high positive value of the path coefficient (β = .690) between Vocabulary and Academic Reading Comprehension was significantly greater than zero as indicated by a one-sample t-test ($t = 9.363$, $p < .001$). The substantial effect size ($R^2 = .476$) indicated the practical significance of this relationship, reflecting that 47.6% of the variance in the linear combination of the IELTS scores was explained by Vocabulary. The high positive value of the path coefficient (β = .763) between Vocabulary and General Reading Comprehension was significantly greater than zero as indicated by a one-sample t-test ($t = 16.059$, $p < .001$). The substantial effect size ($R^2 = .583$) indicated the practical significance of this relationship, reflecting that 58.3% of the variance in the linear combination of the PET scores was explained by Vocabulary.

In conclusion, the results of the modelling using PLS-SEM in Phase 1 provided sufficient evidence to support the stated hypothesis, H2: The receptive vocabulary size of the elementary Emirati learners will correlate with the PET scores. The correlation with the IELTS scores will be strong only for the students with a higher receptive vocabulary size.

Phase 2: Interviews

The results of Phase 2 were interpreted mainly to address RQ3: What are the perceptions of elementary level Emirati learners of English regarding the learning of vocabulary and its relationship to reading comprehension? The results of Phase 2 were also interpreted to enrich the answers to RQ1: To what extent, and in what ways, does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this

element? and RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores? Consequently, the results also include themes that do not necessarily apply directly to answering RQ1 and RQ2 or testing H1 and H2.

The interview transcripts contained a total of 395 statements, including the interventions of the researcher. Table 24 shows that a total of 153 significant statements were extracted from the interview transcripts to identify five primary themes. Each significant statement consisted of verbatim phrases or sentences spoken by a student to identify a named unit of information that could be interpreted to address the research questions. All of the researcher's interventions or irrelevant statements were excluded. Each student's significant statement was coded with the student's level (3 or 4); the student's group (C = Control or E = Experimental); and the student's code name (S1 to S6). The primary themes containing the highest frequencies of statements, and the highest coverage (%), were Theme 1: Reading (55, 36.2%); Theme 2: Vocabulary (44, 28.9%); and Theme 3: Exam (37, 24.33%). Theme 4: Speaking and Listening (9, 5.2%) and Theme 5: Writing (7, 4.0%) had the least coverage. The primary themes were refined into sub-themes, which sub-divided the primary themes into different manifestations of each primary theme. The statements that were aggregated within each theme were tabulated as recommended by Maguire and Delahunt (2017) for presenting the results of a thematic analysis.

Table 24

Primary Themes Extracted from Interview Transcripts

Primary theme	Frequency	Coverage (% of total number of themes)
1: Reading	55	36.2
2: Vocabulary	44	28.9
3: Exam	37	24.3

4: Speaking and Listening	9	5.9
5: Writing	7	4.6

Phase 2. Primary Theme 1: Reading.

A total of 55 significant statements were aggregated into Theme 1: Reading, with five sub-themes. The sub-theme with the highest frequency of statements was Difficulties (24, 43.6%); followed by Practice (18, 32.7%); Text Format (9, 16.4%); and Topic (4, 7.3%). Table 25 presents the statements provided by students expressing their reading difficulties. These statements reflect the many reasons why the students experienced reading difficulties. Table 26 presents the statements provided by students describing how they practised reading in English at home outside of class, including books and newspapers. Two students commented that they read little in English, stating 'Yes, I read book...Yes in Arabic...of course'; and 'Not a lot in English. Not really'.

Table 25

Primary Theme 1: Reading; Sub-theme: Difficulties

Significant statement		Level	Group	Speaker
76	It's hard to know what the question means	4	C	S1
7	Reading, the last part was a little hard.	3	C	S2
9	The reading, last part...It was a little bit difficult because the paragraph was actually full of details and has difficult vocabulary, when I read I'm not sure which one is the answer	3	C	S2
77	We feel difficult level with the reading, especially some part after it is part two and part three, it was a little bit difficult so we hope we did well in last exam.	4	C	S2
109	Yes, sometimes if you don't know how to pronounce the word you feel it difficult to read.	4	C	S3
12	Yes, I think it was difficult especially in the last part.	3	C	S3
48	I think it is - yes.	3	C	S4
26	Reading was hard for me because when I read the text I do not understand a new word. Second and last part is a lot difficult	3	C	S4
28	When we read them my class we don't know what to do and leave them.	3	C	S4

24	I think last part was difficult because you've got more to read and parts you don't understand. I can't read all this in time.	3	C	S3
18	Because it took me a long time.	3	C	S3
139	Because the words are not clear.	3	E	S1
140	The reading... is very difficult	3	E	S2
270	The reading, the work of the reading.	4	E	S2
287	Not that great, because I think about the results. If I do not do well, I will fail to read	4	E	S3
272	It's a problem, the reading ... Some words, we don't know before.	4	E	S3
150	Reading is more difficult for all. Yes.	3	E	S4
156	I feel not happy when I read and I don't understand the words, but I read it all. Maybe I remember, think, or understand.	3	E	S4
275	Because we don't have experience [for typing?] very fast. It's hard for us and very hard too.	4	E	S4
277	it's very hard because there's no time and we see the words the first time we see it.	4	E	S5
279	Yes, and we don't understand the subject	4	E	S5
256	The word's not clear	3	E	S5
290	I feel nervous, because if I don't read the passage twice I will not understand it.	4	E	S4
292	For three texts in one hour with a question, no time.	4	E	S4

Table 26

Primary Theme 1: Reading; Sub-theme: Practice

Significant statement		Level	Group	Speaker
61	Sometimes I read book for example	3	C	S1
107	This is actually because we not read a lot in English, especially in newspapers or books. Usually, we read in Arabic so if we practice reading a newspaper in English or books in English it will become easier than now.	4	C	S1
62	For me sometimes I read newspapers (In Arabic?) Yes	3	C	S2
111	Also, both of them. If we read newspaper more. I feel it will improve our vocabulary and our speed in the reading.	4	C	S3
66	Short stories, graded readers	3	C	S3
50	Because we do not practise at home and during class we can't read it faster.	3	C	S4
68	I like the videos and if I like the story I can read that story	3	C	S4
90	For me, I feel like I have to read more than the regular reading I do. I think I can just read and learn from the college so I have to read more at home or in my normal life, so my vocabulary became better than before.	4	C	S5
376	One or two books	4	E	S1

369	Yes, I read books	4	E	S2
373	Yes in Arabic...of course.	4	E	S2
375	In a year, one or two books	4	E	S2
365	Yes, I read newspapers.	4	E	S3
381	Not a lot in English. Not really.	4	E	S6
377	Actually, I read story books and of science.	4	E	S6
228	Yes. It is easy to read fast after practice	3	E	S2
231	Yes. Of course. [Speed reading]	3	E	S5
229	Yes. It's easier because every day and every today, do the work in speed reading. Now it's easy, little, little.	3	E	S5

Table 27 presents the statements provided by two students in the control group and four students in the experimental group describing how they handled the text format (i.e., reading in English from left to right, as opposed to in Arabic, for right to left). All of the students suggested that they had no reading difficulties associated with the text format,

Table 27

Primary Theme 1: Reading; Sub-theme: Text Format

Significant statement		Level	Group	Speaker
57	I think it is no problem for the student but I think more student read Arabic than in English.	3	C	S1
117	I think it's not hard for to recognise a word the English starts, the layout I mean in the essays and these things, because you know it from left to right. You cannot start from the other way because it's obviously wrong.	4	C	S2
235	No, not a problem.	3	E	S1
237	Because in the school I already learn Arab this and right and English left.	3	E	S1
355	No problem in reading from here and there, from the right and left, and I can understand them easily.	4	E	S2
361	Not any problem.	4	E	S3
239	It's not a problem.	3	E	S4
243	Not a problem.	3	E	S4
353	No, it's easy.	4	E	S4

Table 28 presents four statements provided by three students in the experimental group suggesting that the topic was not necessarily a source of reading difficulty.

Table 28

Primary Theme 1: Reading; Sub-theme: Topic

Significant statement		Level	Group	Speaker
384	If you know it (the topic).	4	E	S1
383	When I read the topic, the title, the reading, I feel relaxed and I can read regardless of the topic.	4	E	S3
385	Yeah. I can imagine what I will read after this topic. And I can read it easily after that [like?] I used to. But if the subject [inaudible] is confusing.	4	E	S3
391	Yes. If the vocabulary is easy, it doesn't matter about the topic so much	4	E	S5

Phase 2. Primary Theme 2: Vocabulary

A total of 44 significant statements were aggregated into Theme 2: Vocabulary, with seven sub-themes. The sub-theme with the highest frequency of statements was Dictionaries (15, 34.1%) followed by Word Cards and translation (9, 20.5%); Difficulties (7, 15.9%); Speed-reading (6, 13.6%); and Teacher (6.8%). Table 29 presents seven statements provided by one student in the control group and six students the experimental group expressing the difficulties they experienced with learning vocabulary.

Table 29

Primary Theme 2: Vocabulary; Sub-theme: Difficulties

Significant statement		Level	Group	Speaker
89	When we see a vocabulary which we cannot understand about it we blame ourselves because, if we study little more than normal study which we study, we will feel it easier than before.	4	C	S1
160	The vocab is very difficult and new words I don't see before.	3	E	S2
142	Because there is many vocabulary.	3	E	S2
154	Problems with the vocabulary. Yes.	3	E	S4
204	To learn and know, and every day I need new vocabulary to learn.	3	E	S4
159	The vocab is not clear. I don't understand.	3	E	S5
248	Instead for paragraph three to say the word and to get a new-- another vocabulary. It's very difficult.	3	E	S5

Table 30 presents seven statements provided by one student in the control group and three students the experimental group expressing their view that online dictionaries (e.g., Wordpower) are useful for learning English vocabulary. Table 31 presents seven statements provided by three students in the control group expressing their views about English dictionaries (e.g. Oxford English). Two students did not like English dictionaries because it was 'Slow to find words' and 'we don't understand it by English translation, we want it in Arabic'.

Table 30

Primary Theme 2: Vocabulary; Sub-theme: Online Dictionaries

Significant statement		Level	Group	Speaker
326	I liked it because I know the meaning of the word in Arabic. And I understand the meaning of every word. If I hear it only in English I don't understand.	4	E	S2
323	Save on the time.	4	E	S3
395	When you, the teacher find from dictionary- have many meaning and we don't know every meaning.	4	E	S5
327	Also, it tells if the word has different meaning, it shows for	4	E	S5

	how many meaning it has.			
324	Saving on time, right, and I will know what the real meaning of the word. And I can use the word when I want, so I know the meaning of the word, and I can use it and sometimes I [inaudible].	4	E	S5
32	Because it has a lot of different meanings and helps us to learn and have meaning for one word.	3	C	S2
34	Yes, it's better than other dictionary - it's on computer.	3	C	S2
44	It is - It is good. I think the BB Vista is good for student because you here the sound of the word.	3	C	S2
46	It helps practice	3	C	S3

Table 31

Primary Theme 2: Vocabulary; Sub-theme: English Dictionaries

Significant statement		Level	Group	Speaker
99	Yes, or the Oxford dictionary also.	4	C	S2
101	Yes, we use it every day because it's really important that if you didn't know the words you can just know it by yourself by just look for it in the dictionary without the help from the teacher	4	C	S2
36	Slow to find words	3	C	S3
103	Because we don't understand it by English translation, we want it in Arabic because it's much easier than English. Sometimes when they write it in English transfer is very hard some words, I must look at it another time to see what the word mean.	4	C	S4

Table 32 presents six statements provided by four students in the experimental group expressing their views that they identified words when speed reading. Table 33 presents four statements provided by students in the experimental group who used Words to Know.

Table 32

Primary Theme 2: Vocabulary; Sub-theme: Speed-reading

Significant statement		Level	Group	Speaker
321	It's easier for us.	4	E	S2
320	Was easy.	4	E	S3
315	Yes. I notice those words in the speed-reading.	4	E	S4
318	Yes. Speed-reading is easy.	4	E	S4
316	I notice some of them in the speed-reading	4	E	S5
389	Because the vocabulary is easy (in speed reading)	4	E	S5

Table 33

Primary Theme 2: Vocabulary; Sub-theme: Words to Know

Significant statement		Level	Group	Speaker
95	Use Words to Know	4	E	All
187	Speed read on Words to Know	3	E	S1
195	Yes, I did learn how to spell the words	3	E	S1
185	Words to Know is very good because it translates Arabic into English	3	E	S1

Table 34 presents three statements provided by two students in the control group who appreciated talking with the teacher to help explain the meaning of words.

Table 34

Primary Theme 2: Vocabulary; Sub-theme: Teacher

Significant statement		Level	Group	Speaker
40	When you find a new word and write it down, the teacher can explain meaning for you.	3	C	S1
125	A lot of new vocabulary we learnt because of many things – like listening to the teachers when they talk, they explain for us what does this word mean.	4	C	S1
92	When we talk with our teachers we know how to pronounce the word and there's a new word that they explain it to us if we don't know the meaning of it they will tell us what its mean and it will be clear and we will, can use it more.	4	C	S5

Table 35 presents nine statements provided all members of the experimental group suggested that the use of Word Cards and Translation was good because it helped them to learn and understand the words quickly.

Table 35

Primary Theme 2: Vocabulary; Sub-theme: Word Cards and Translation

Significant statement		Level	Group	Speaker
183	Yes, it's very quick for me, but because I understand the word.	3	E	S1
166	It's good for learning.	3	E	S1
170	Because it's easy for learn.	3	E	S1
307	I think it's, the words on the cards, it's good way but I think...If you can't read the words, you can't understand	4	E	S2
306	For me it's a new method for learning. Because last year when I studied a word I just print it in a paper and just memorise it. But now it's in a card. I can practice a word or exercises and I can share it with my partner also.	4	E	S3
171	We understand quickly because in the card, the word translates it in Arabic and we see the word in Arabic and English so--	3	E	S3
173	Word cards were a good thing. Yes	3	E	S4
393	I think that the cards, the cards they are good. In Arab have a different meaning.	4	E	S5
197	I will learn to understand the word and meaning for this.	3	E	S5

Phase 2. Primary Theme 3: Exam

A total of 37 significant statements were aggregated into Theme 3: Exam, with five sub-themes. The sub-theme with the highest frequency of statements was IELTS vs. PET (15, 40.5%) followed by Difficult (11, 29.7%); Easy (6, 16.2%); and Time (5, 13.5%). Table 36 presents nine statements expressing the view that the exam was difficult.

Table 36

Primary Theme 3: Exam; Sub-theme: Difficult

Significant statement		Level	Group	Speaker
14	A lot of hard parts	3	C	S2
7	Reading, the last part was a little hard.	3	C	S2
	I think last part was difficult because you've got more to read and parts you don't understand.	3	C	S3
24	I can't read all this in time.			
9	The reading, last part, It was a little bit difficult because the paragraph was actually full of details and has difficult vocabulary, when I read I can't not sure which ones the answer	3	C	S2
266	It was very hard for us. Very tough.	4	E	S2
268	Actually, of course we have a time for studying but it's hard.	4	E	S2
274	Yes, both are hard	4	E	S3
295	Yes, but I cannot, I think about something else.	4	E	S5
296	I feel confused because half the time I only understood one, and I still don't solve passage two and three.	4	E	S6

Table 37 presents five statements expressing the view that parts of the exam were easy and other parts were difficult. Table 38 presents 15 statements expressing the view that the IELTS exam was the most difficult, mainly because 'IELTS has more difficult words'. Table 39 presents five statements expressing the view that there was not enough time to do the exam.

Table 37

Primary Theme 3: Exam; Sub-theme: Easy

Significant statement		Level	Group	Speaker
3	I think the exam was a little bit easy and - the reading was easy	3	C	S1
76	Some of it was easy	4	C	S1
26	For me the first part was easy	3	C	S4
135	The listening exam is very easy.	3	E	S1
148	I think the PET exam is easy.	3	E	S3
140	I think the PET exam is in the middle.	3	E	S2

Table 38

Primary Theme3: Exam; Sub-theme IELTS vs, PET

Significant statement		Level	Group	Speaker
83	Actually, IELTS is harder than PET.	4	C	S1
87	PET is easier, you can think about it and the vocab is good for beginners and the others.	4	C	S1
85	IELTS, because it includes a lot of vocabulary we don't know it.	4	C	S1
79	I think that because part two has more vocabulary and it's a little bit taller (longer) I think, it has more words to read and I think the main purpose because it was about business, and in business there are so many words about it that are hard to know or understand.	4	C	S2
80	Part one had more vocabulary we know and the part two some of it difficult, some of it easy.	4	C	S3
81	There is some word in part one that we know and we often read it but part two and part three there is a new vocabulary that we don't know; new word that is difficult.	4	C	S4
341	Many words. Many words	4	E	S2
282	The IELTS exam is harder than the [PET?] exam.	4	E	S2
350	Yeah. PET was an easier exam.	4	E	S3
347	The IELTS had the most difficult for vocabulary.	4	E	S3
339	Yes. Lot of unknown words in IELTS	4	E	S4
351	It's easier than the IELTS.	4	E	S5
348	IELTS has more difficult words	4	E	S5
340	Many, many unknown words in IELTS	4	E	S5
343	For me, 70% (unknown words in IELTS)	4	E	S6

Table *39*

Primary Theme 3: Exam; Sub-theme: Time

Significant statement		Level	Group	Speaker
18	It took me a long time.	3	C	S3
219	The time that is problem in the PET exam.	3	E	S1
285	I feel funny really, because I don't have time, and I'm not reading fast. So, I cannot concentrate, and I think I will miss the words or miss something. And I feel funny.	4	E	S2
223	You don't have the time, enough time...if there are new words, a new vocabulary, on the bar graph.	3	E	S3
302	It was like a guess. There was no time	4	E	S6

Phase 2. Primary Theme 4: Speaking and Listening

Table 40 presents the seven statements that were aggregated into Theme 4: Speaking and Listening.

Table 40

Primary Theme 4: Speaking and Listening

	Significant statement	Level	Group	Speaker	Sub-theme
140	The listening is very difficult	3	E	S2	Difficult listening
146	Because he is speaking very fast.	3	E	S2	Difficult listening
115	Could hear words on Words to Know	4	E	All	Easy listening
148	The listening is not clear, but easy.	3	E	S3	Easy listening
189	Yes, very clear.	3	E	S1	Easy listening
191	Yes, it is. It is useful.	3	E	S1	Easy listening
193	When I listen.... I understand what it means.	3	E	S1	Easy listening
113	Need to know how to say it as well as to read it	4	C	S3	Saying words
313	Yes. Saying it later on Blackboard help me.	4	E	S2	Saying words

One student in the experimental group suggested that listening to words, (e.g. in exams), was difficult. All students in the experimental group said that they could listen to Words to Know. One student in the control group and one in the experimental group suggested that speaking the words was useful.

Phase 2. Primary Theme 5: Writing

Table 41 presents the 7 statements that were aggregated into Theme 5: Writing

Table 41

Primary Theme 5: Writing

Significant statement		Level	Group	Speaker	Sub-theme
3	The writing was very hard.	3	C	S1	Difficult
74	I write news words in notebook with notes	3	C	S1	Notebook
252	Yes, I first write the paragraph...and another paragraph...to put the word there and that's difficult	3	E	S5	Difficult
257	I don't understand the word. Maybe it says another thing and so I write the wrong word	3	E	S2	Difficult
275	We don't have experience for typing very fast. It's hard for us and…very hard too	4	E	S4	Difficult
203	Writing helps to learn and to remember words.	3	E	S5	Helps to learn
199	I write the words on Blackboard	3	E	S5	Blackboard

One student in the control group and three in the experimental group expressed their difficulties with writing words. One student in the control group suggested writing new words in a notebook was useful. One student in the experimental group suggested that writing the words on Blackboard helped him to learn and remember words.

CHAPTER 5: Discussion and Conclusions

Introduction

Chapter 5 contains the following five sections. The first section presents a summary and synthesis of the findings obtained to address RQ1, RQ2, and RQ3, and test H1, including the outcomes of the triangulation, and the researcher's interpretation of the results in the context of the literature. The second section discusses the implications of the findings and the contributions of the study with respect to (a) theories associated with the learning of vocabulary; (b) the use of a mixed methods methodology; and (c) policies and practices in educational settings. The third section discusses the limitations of the study. The fourth section considers the need for future research, in the light of the limitations of the current research. The chapter ends with some final remarks.

Summary and Synthesis of the Findings

This section contains three sub-headings, as follows: RQ1: Decontextualised vocabulary study and gain in receptive vocabulary; RQ2: Vocabulary size and reading comprehension; and RQ3: Perceptions

RQ1: Decontextualized vocabulary and gain in receptive vocabulary

Empirical evidence was provided to address RQ1: To what extent, and in what ways, does decontextualised vocabulary study, using word cards and translation, contribute to greater gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element? The researcher addressed the problems posed in RQ1 by providing empirical quantitative and qualitative data to test H1: Decontextualised vocabulary

study, using word cards and translation, may contribute toward a more rapid gain and a greater amount of receptive vocabulary among elementary level Emirati learners of English than a similar teaching programme lacking this element.

The primary strength of the experiment involving the analysis of test scores for the VST, VLT, and AWT tests was that it was continued for an academic year, over two semesters, with multiple repeated measures to test for vocabulary size. Therefore, this experiment was unlike previous experiments to test the use of word cards, which were only continued for a few weeks (e.g. Kuo & Ho, 2012; Nation, 2001; Ramachandran & Rahim, 2004). Another strength of the quantitative data was that the total sample size of 37 Arab learners of English in each group was adequate to provide sufficient power to generate statistically significant results at $p < .05$, based on the results of a power analysis, assuming a moderate effect size. Therefore, the statistical analysis was not underpowered, because, by definition, underpowered means that the results of a statistical tests are falsely declared to be not significant, when in fact, they should be significant (Maxwell & Kelly, 2011).

The results of repeated measures ANOVA on the vocabulary tests scores (VLT, VST, and AWT) indicated that the students in the experimental group appeared to have learned vocabulary quicker using the word card and translation approach than the control group, who did not use that approach. The statistically significant time x group interaction meant that the pattern of change in the vocabulary gain scores over time was not the same in the experimental group as in the control group. The problem with comparing the impact of an intervention between groups over time using ANOVA, is that it is very difficult to evaluate statistically the strength of the between-subject effect (i.e., the amount of gain in vocabulary) if this effect is confounded by a

significant within-subject interaction of time x group (i.e., the speed of gain in vocabulary). The interpretation of the results of the repeated measures ANOVA were confounded because the strength of the impact of the intervention on the between-subject effect was not exactly equivalent for each group (Crowder & Hand, 1990; Frison & Pocock, 1992; Hair et al., 2010). This may explain why (a) the group had no significant effect on the amount of gain in the VLT scores (indicated by $p > .05$, with negligible effect sizes). Although the between-subject effect for the gain in the amount of the VST scores was significantly different ($p < .05$) indicating a difference between the amount of gain in vocabulary of the experimental and control groups, the effect size was small. The between-subject effect reported for the gain in the amount of AWL scores was not significantly different with respect to the group with a negligible effect size. Consequently, the main conclusion based on the results of the repeated measures ANOVA was that decontextualised vocabulary study, using word cards and translation, may contribute toward a more rapid gain of receptive vocabulary among elementary level Emirati learners of English, than a similar teaching programme lacking this element

Qualitative evidence to further address RQ1 and support H1 was provided by triangulation. The results of the survey and experiment were corroborated by the interview responses as outlined in Table 42. The frequencies of the responses to the interview questions are reported in Table 42, because a high frequency of responses regarding a certain issue implies that the participants may perceive that that issue is very important. Conversely, a low frequency of responses implies that the participants may perceive that the issue has less importance (Yin, 2014). A very high proportion (nine of eleven) of the interviewed members of the experimental

group suggested that word cards and translation were good because they helped them quickly to learn and understand the words, consistent with the results of the repeated measures ANOVA.

It is important not to take the views of the experimental students on word cards in isolation. Table 25 shows that students from both control and experimental groups were very clear that the major difficulty they had when reading, especially in examinations, was caused by unknown vocabulary. These excerpts illustrate the point: ES1. Because the words are not clear; ES3. It's a problem, the reading ... Some words, we don't know before; CS4. … when I read the text I do not understand a new word; CS2. … full of details and has difficult vocabulary. These students had identified that expanding their vocabulary size was the key issue if they were to read better. Therefore, using an approach to learning vocabulary which they found 'easy', 'quick' and could understand clearly improved their belief in themselves as well as growing their vocabulary size and improving their reading.

Table 42

Triangulation of Results: Survey, Experiment (Phase 1); Interviews (Phase 2)

Question	Phase 1: Survey	Phase 1: Experiment	Phase 2: Interviews
To what extent?	The responses to the survey questions indicated that the members of the experimental group generally agreed that learning vocabulary using Word Cards was very useful (M = 4.31, SD = 0.93)	Statistically significant ($p < .001$) time x group interaction for the VLT, VST, and AWL test scores. The effect sizes indicated the practical significance of the strength of the interactions ($\eta^2 = 0.43$, 0.78, and 0.16 respectively)	Nine statements were provided by members of the experimental group concerning the use of word cards and translation.
In what ways?		The VLT, VST, and AWL scores in the experimental group increased more rapidly with time than in the control group.	Students in both the control and experimental groups expressed the difficulties that they experienced with learning vocabulary. Nine members of the experimental group suggested that word cards and translation were good because they helped them quickly to learn and understand the words.

The conclusion that decontextualised vocabulary study, using word cards and translation, resulted in a more rapid gain in receptive vocabulary for elementary level Emirati learners of English was consistent with the literature. Other researchers (Nation, 2001; Ramachandran & Rahim, 2004; Kuo & Ho, 2012) similarly supported the use of word cards and translation for the rapid learning of vocabulary among elementary level learners of English as a foreign language.

The analysis of the VLT, VST, and AWT scores also appeared to support the conclusions of other researchers that significant differences exist between the vocabularies of low and high-levels of English language learners (Horst, 2000; Zahar, Cobb & Spada, 2001).

RQ2: Vocabulary size and reading comprehension

Empirical evidence was provided to address RQ2: What is the relationship between receptive vocabulary size and reading comprehension scores? The researcher addressed the problems posed in RQ1 by providing empirical quantitative and qualitative data to test H2: The receptive vocabulary size of the elementary Emirati learners correlate with the PET reading scores. In contrast, the IELTS reading scores correlate only with the receptive vocabulary size of those participants who exhibited the greatest receptive vocabulary gains.

The results of the PLS-SEM analysis of the vocabulary tests (VLT, VST, AWT scores) and reading comprehension tests (IELTS and PET scores) provided the statistical evidence to concluded that the receptive vocabulary size correlated with reading comprehension scores, as measured by IELTS and PET reading tests. The results of factor analysis confirmed that the two reading comprehension tests measured two different constructs, probably because the IELTS and PET scores tapped different skills. The descriptive statistics showed that the mean percentage score of both the experimental and control groups of students was higher on the PET tests than on the IELTS tests. Triangulation indicated that these statistics were corroborated by the qualitative data, as outlined in Table 43. All of the students in the control and experimental groups expressed the view that the IELTS exam was the most difficult, mainly because ‘IELTS has more difficult words’. The interview data on speed-reading in Table 32 supports the students’ perception that their difficulties in reading are primarily caused by a lack of vocabulary

knowledge. The six statements from four experimental group students emphasise that they found speed-reading ‘easy’. We should remember that the 20 speed-reading course texts were all written using only the 1000 most frequent English words. It sees apparent that it was ‘easy’ because by the stage speed-reading was introduced the students had achieved vocabulary sizes of 1000 or more words. In addition, two of the students also commented that they had noticed the words they had previously learnt from word cards. The pattern that emerges from both the quantitative and qualitative data is clear. The students found speed-reading easy because of low vocabulary demands; they could cope quite well with PET even though the vocabulary load was still somewhat higher than their vocabulary size; they had major difficulty with the IELTS vocabulary, presumably because many of the words were well outside of their range. The evidence underlines that the IELTS reading module, demanded by the institution involved, is not appropriate for these students, Furthermore, the evidence strongly suggests that these elementary students will only enjoy success in reading if the tests used have vocabulary loads generally in line with the students’ vocabulary size, In other words, it is not that these students cannot read per se but that they have great difficulty reading words they have not learnt. In the light of self-efficacy theory, discussed in the literature review, it seems probable that if only IELTS reading texts had been used with these students, their belief in their ability to read could have been badly shaken. Fortunately, the incorporation of speed-reading and PET texts mitigated this possibility.

The possibility that the PET and IELTS test tapped different skills was also indicated by the results PLS-SEM analysis showing that the correlation between the PET scores and vocabulary size was stronger than the correlation between the IELTS scores and the vocabulary size. The substantial effect size (R^2 =.476 for the IELTS scores and R2 = .583 for the PET scores

indicated the practical significance of these results (Ferguson, 2009; Fraenkel & Wallen, 2011). These findings are consistent with previous studies concluding that the IELTS test is very difficult for Arab learners of English in the UAE (IELTS Annual Reviews, 2006, 2007; Cambridge ESOL research notes, 2010). As we have seen in the 'instruments' section of the methodology chapter, the IELTS reading module was designed to be a test for candidates who wish to study in English medium higher education institutions (IELTS Handbook, 2007). As such, the reading texts used in the reading test, are at an approximately equivalent level of vocabulary load and grammatical difficulty to those read in university undergraduate programme (Green, Unaldi & Weir, 2009). Therefore, whilst IELTS appears to be an appropriate test for its target audience, it is too difficult for elementary level English students and is unable to accurately measure these students reading skills. The present situation may not be ideal for Arab students who wish to pursue their studies in an English medium institution based on the results achieved using IELTS tests.

Table 43

Triangulation of Results: Modelling (Phase 1); Interviews (Phase 2)

Question	Phase 1: Modelling	Phase 2: Interviews
To what extent?	The substantial effect size (R^2 = .476) indicated that 47.6% of the variance in the linear combination of the IELTS scores was explained by Vocabulary. The substantial effect size (R^2 = .583) indicated that 58.3% of the variance in the linear combination of the PET scores was explained by Vocabulary.	Nine students in the control and experimental groups expressed the difficulties that they experienced with the exams**.**
In what ways?	The receptive vocabulary size of the elementary Emirati learners was more strongly correlated with the PET scores than the IELT scores.	All students in the control and experimental groups expressed the view that the IELTS exam was the most difficult, mainly because "IELTS has more difficult words"

RQ3: Perceptions

Empirical qualitative evidence was provided to address RQ3: What are the perceptions of elementary level Emirati learners of English regarding the learning of vocabulary and its relationship to reading comprehension? In addition to the outcomes of triangulation, outlined in Tables 42 and 43, the responses to the interview questions provided additional information, to enrich and explain the answers to RQ1 and RQ2. The information presented in Tables 42 and 43 indicated that the outcomes of the triangulation were successful. The findings of the qualitative studies appeared to be consistent with, converge with, corroborate, and helped to confirm the findings of the quantitative study.

In the case of RQ3, the argument that their reading improved is exemplified not only by their test results but also by the excerpts from the interviews in tables 37 and 38. These clearly indicate that the students began to find the PET reading examination 'easy', although their perception of the IELTS reading examination was that it was 'difficult' The reasons given overwhelmingly concern difficulties with unknown vocabulary in IELTS but not in PET. As was discussed in the PET section of Phase 1: Instruments, the vocabulary load of PET reading is limited to 2708 words of the most frequent English words (Street & Ingham, 2007). IELTS by contrast, although controlled, has comparable vocabulary to what might be read in the first year of an undergraduate degree (Weir et al, 2009). These findings from the interviews seem to confirm that the students had begun to see they might become successful readers and that they had clearly identified a key issue in doing this; expanding their vocabulary knowledge. It was also evident from the interviews that the experimental students were now convinced that they knew how to achieve the required increase in vocabulary through the word card approach (table 35). These findings are theoretically supported by self-efficacy theory as discussed in the literature review. If students discover how to succeed and begin to succeed, their self-belief grows and they become more likely to be successful.

The thematic analysis of the interview responses explained why some elementary level Emirati learners of English perceived that they achieved better test scores for vocabulary and reading comprehension than others was that they learnt English vocabulary and reading comprehension outside of class, without the use of word cards and translation. For example, seven statements were provided by the interviewed students describing how they practised reading in English at home outside of class. However, reading outside of college is not the norm

for male Emirati students. Cultural factors such as family responsibility limit the time of any studying once the college day has finished. Males are expected to chaperone their female relatives and conduct other family related business (Harb & El-Shaarawi, 2007). Conversely, the popularity and success of word cards is likely to have been enhanced because Emirati culture tends to view learning as something achieved, not by reflection, but as something given by an expert, in this case the teacher (Richardson, 2004).

The results of the survey also indicated that the experimental group perceived that online learning was the best way to learn English vocabulary. The two online learning programmes embedded in the experimental group's treatment is described in detail in the 'procedure' section of the methodology chapter. It seems likely that the students valued this type of learning because the instructions for use were made clear at the outset, the design meant that both students and teachers could track progress as students worked through word sets and the 'Words to Know' programme included game like activities where the students competed against themselves to increase speed or against others in activities such as 'Listen-Match' and Spelling Web'. These qualities in online work have led to increased motivation and student success in other studies as discussed in the 'online resources' section of the Literature Review chapter. Additionally, as students became increasingly more successful their motivation presumably grew as claimed by self-efficacy theory. Table 33 highlights that the interviews provided further evidence of experimental group students favouring online learning, specifically 'Words to Know'. The reasons given include the use of translation and help in learning to spell. The use of translation and online learning are recurring themes. Table 11 survey results show that both the control and experimental groups strongly favoured online bilingual dictionaries over book form English only

dictionaries. The interview data in table 30 clarifies why the students preferred the online bilingual dictionaries; that translation made the word meaning clear and that these dictionaries gave quick access to information.

The thematic analysis of the interview responses also revealed that the students perceived that differences in their levels of speaking, listening, and writing English may be another reason why they may have achieved better test scores for vocabulary and reading comprehension than others. Two themes, called "Speaking and Listening" and "Writing" were extracted from the interview responses. All students in the experimental group said that liked listening to Words to Know. One student in the control group and one in the experimental group suggested that speaking the words was useful. One student in the control group suggested writing new words in a notebook was useful. One student in the experimental group suggested that writing the words on Blackboard helped him to learn and remember words. These themes are consistent with the benefits of a multi-sensory approach to learning English as a foreign language. Multi-sensory is a term used to refer to any learning activity that combines multiple senses (e.g., using a mixture of visual, auditory and other sensory modalities). Jubran (2012) concluded that many students were more engaged in learning English when they were given a chance to use two or more senses.

Limitations of the Findings

The quantitative phase of this study was subject to threats to internal validity because factors other than the effects of the prescribed educational intervention may have been responsible for the variance in the test scores. These variables could include the amount of practice in English vocabulary and reading comprehensions that the students experienced at home, outside of the formal classroom sessions, and the differences in vocabulary knowledge at

the baseline. An independent samples *t*-test indicated that the mean VST score of the experimental group at Time 1 and the mean VST score of the control group at Time 1 were significantly different. Therefore, the assumption was violated that the vocabulary size of the two groups should be equivalent before the intervention. The groups had been randomly assigned on entry to the programme as described in the 'population' section of the methodology chapter. However, it should also be noted that random assignment of participants into groups does not necessarily ensure that the groups are identical at the start of an experiment with respect to the variables under investigation ((Price, Jhangiani and I-Chant, 2018)). The control group were only considered to be stronger because the p-value of the t-test on the vocabulary mean scores was < .05; however, when we look closer at the actual vocabulary size it can be argued that the difference would not be practically significant in this study. In fact, in terms of CEPA, the strength of the groups was similar at the outset. The initial testing of CEPA means, after random group assignment, returned a mean of 159.54 for the experimental group students and 161.68 for the control group students. However, the mean testing for vocabulary size at the pre-treatment stage (week2) produced means of 4.324 for the all experimental group and 5.648 for all the control group students. The mean of the experimental group equates to approximately 430 words whilst that of the control group is approximately 560 words. As we can see, the control group was found to be stronger in terms of vocabulary. This might be seen as a mitigating factor in terms of group comparability given the purpose of this study. The confounding variable in this case would not logically be seen to confuse the interpretation of results. That the experimental group was weaker in vocabulary on entry, but stronger on completion of the study, might instead

be seen as highlighting the efficacy of the intervention. We should also note that these vocabulary sizes identify all of the students in both groups as elementary level.

The quantitative phase of this study also had weaknesses due the common misconception among quantitative educational researchers called the ecological fallacy, defined as the assumption that each individual in a group behaves in exactly the same way as the mean score or some other statistic computed to summarise the whole group (May et al., 2003). The ecological fallacy is a source of bias in educational research, because students are often classified into groups, and mean scores are analysed to summarize and compare the academic achievements of each group. The ecological fallacy was implicated in this study, because of the large variance in the test scores. The large variance was due the scores of some the students being much lower than the mean scores, whereas the test scores of other students were much greater than the mean test scores; however, none of the students actually achieved the mean test scores, even though the mean scores were used as a basis for comparing the groups. An example is given in Figure 9, displaying the frequency distribution histogram of the positive VST gain scores (i.e., the final score minus the first score) among the 37 members of the experimental group.

The mean gain score for the VST was $M = 6.45$; however, 17 students achieved VST scores lower than 6.45, no students achieved the mean score of 6.45, and 30 students achieved VST scores greater than 6.45. The implications are that the impact of the decontextualized vocabulary study, using word cards and translation, appeared to be much better for some students (with VST scores higher than the mean) than for other students (with VST scores lower than the mean) supporting the ecological fallacy. Not all of the students in the experimental group gained

the same benefit in learning vocabulary from using the word card approach, supporting the concept of the ecological fallacy.

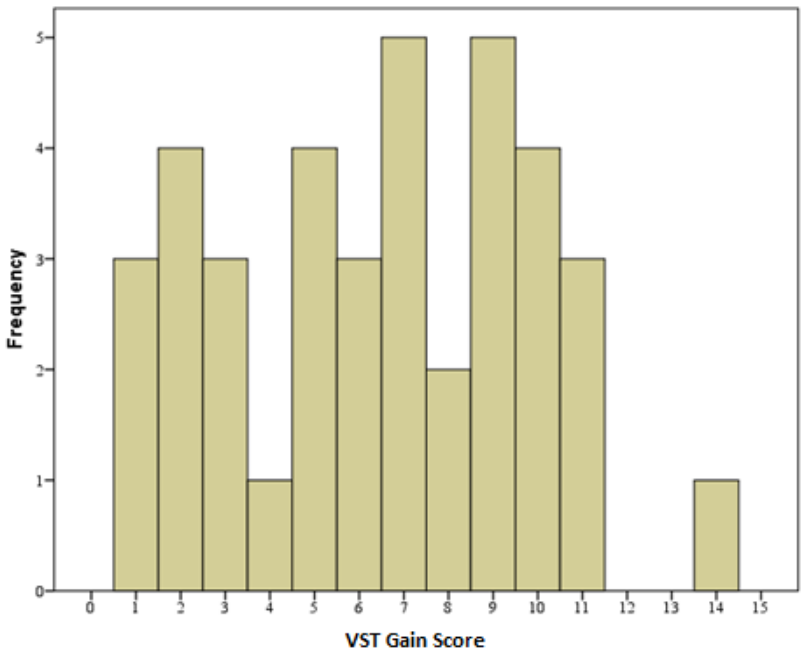

Figure 9. Frequency distribution histogram of VST gain scores

The lack of information about the impact of the prescribed intervention on the academic achievement of each individual student who participated in the current study was due to the restriction of using ANOVA. The statistics applied only to generalizing the results to the two groups of students, and not to each individual student. Consequently, ANOVA did not identify which individual students had difficulty with decontextualized vocabulary study using word cards and translation, which individual students benefited most from the intervention, or why some student benefited more from the intervention than others.

This study may also be seen limited in the form-meaning vocabulary focus. As discussed on page 26, there are many elements to learning a word and the necessary focus in this study

given the level of the students was form-meaning. This can be described as the first step of the vocabulary knowledge journey. How this focus might be widened is outlined in the 'future research' section below.

A further possible limitation was the narrowness of the population involved. It is true that the participants were all male elementary level Emirati students from one college in the UAE. In truth, it was the struggle of this population which first inspired the study. However, despite this their problems with vocabulary and reading are very similar to those of other populations. Their lack of belief in themselves is mirrored by other groups of language learners that are not experiencing success, How the findings of this studied might be applicable to other populations is discussed in detail in the following 'implications' section.

The impact of implementing decontextualized vocabulary study, using word cards and translation on the ability of each elementary level Emirati learner of English to attain success in the vocabulary and reading comprehension tests was not a simple and direct relationship. A possible criticism is that the effect of the treatment could be indirectly moderated by intervening factors. For example, Farrington, Roderick, Allensworth, Nagoaka, et al. (2012) suggested that three non-cognitive factors, specifically "mind-set,"; "perseverance" and "behaviour" intervene to determine if a high-level of academic achievement is the outcome of a prescribed education intervention, as illustrated by the flow diagram in Figure 10

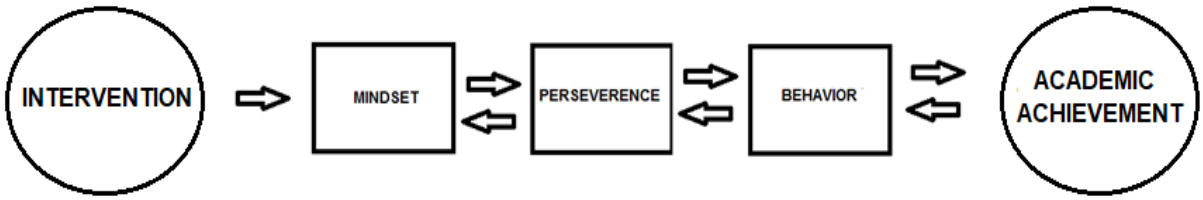

Figure 10. The impact of an educational intervention on academic achievement

If an individual student reacts to a prescribed educational intervention with a positive mind-set (e.g., manifested by optimism, self-confidence, high self-esteem, and expectation of success) then that student will become motivated to persevere with the intervention, resulting in effective learning behaviour, and ultimately leading to good academic performance. Conversely, if an individual student reacts to a prescribed educational intervention with a negative mind-set (e.g., manifested by pessimism, lack of self-confidence, low self-esteem, and expectation of failure) then the student's motivation to persevere will be stifled, effective learning behaviour will be undermined, and the ultimate outcome will be poor academic achievement. There is also a feedback relationship. Good academic achievement strengthens a positive mind-set, increases perseverance, and reinforces effective behaviour. That other factors may have influenced the participants' performance might be seen as a limitation. However, although it was not possible to identify all of the factors leading to good academic performance in the vocabulary and reading comprehension tests, the data on the participants' perceptions points to the word card treatment as the catalyst for the experimental group elementary Emirati learners of English increased belief in their vocabulary learning capabilities. It could then be argued that the vocabulary treatment was responsible for their improved vocabulary size and this success led to increased self-efficacy

(Bandura, 1997) as described by in the Vocabulary Learning Strategies and Self-Efficacy Theory section of the Literature Review. The same section highlights that students' with decreased belief lose the motivation to study and that this lowers academic performance (Hamada, 2014). In contrast, the literature supports the view that students' who are provided with the means to be sure they have understood and are given clear teacher feedback, are likely to improve their performance and gain motivation and belief (Rauber & Gill, 2004).

Accordingly, in the current study (Table 34), two of the interviewed students reported that they appreciated the sensitivity of the teacher in the classroom, for example through the teacher actively helping them to understand the meaning of English words. Therefore, the sensitivity of the teacher may have been a factor that stimulated at least some elementary Emirati learners of English to achieve better performance than others in the current study. By personally supporting the students to improve their individual mind-sets, the teacher may have helped the learners to improve their perseverance, behaviour, and academic achievement. However, we should note that the students in question were members of the control group. As this group were not using word cards or online activities for learning vocabulary it was incumbent on the teachers to explain aspects of the words. In contrast, once the teacher had set up the word card treatment and online activities and students had become accustomed to these, the key aspects of word meaning, pronunciation and spelling were repeatedly covered through the treatment. Also, working with two groups one of 19 and the other of 18 students in the experimental groups allowed the teacher to provide individualized instruction and clarity regarding the use of word cards and translation. When using the word cards the students worked in pairs, or pairs and one group of three in the case of the group of 19, and tested their partner(s) with the word cards. This

approach allowed the teacher to move around the classroom and give attention to the students and advice where needed. However, it should be noted that the control groups worked with very experienced and successful teachers in groups of the same size as the experimental groups. They were also made completely aware of what was required of them and given individual attention when needed.

Implications of the Findings

The implications of the findings of this study for theory, in the specific context of the learning of vocabulary, are difficult to assess, because, within linguistics, word knowledge has only lately become a serious candidate for theorizing and model building (Richards, 1976; Grettell, 1981; Nuttall, 1982). The results, however, do support Hulstijn (2001) suggestion that in L2 teaching theory, intentional vocabulary learning employs activities and strategies that have the specific objective of memorizing vocabulary. Furthermore, the results of this study support self-efficacy theory (Bandura, 1997), because this theory predicts that the goals of elementary level Emirati learners of English to rapidly increase their vocabulary size should be related to the tools (e.g., word cards and translation) that they believe are effective when learning vocabulary. Continual failure to achieve high scores in English vocabulary and reading comprehension tests, may lead learners to believe that their goals cannot be achieved. Underpinned by self-efficacy theory, however, the discovery that elementary level Emirati learners of English language appear to respond rapidly to non-contextual methods of learning vocabulary using word cards appears to carry great importance. Because self-efficacy theory is considered to be a critically important contribution to the study of academic achievement, motivation, and learning, the results of this

study provide a foundation for the development of educational policies and classroom practices (Artino, 2012).

The implications of the results of this study for methodology are that the strengths of the sequential explanatory mixed methods design defined by Creswell (2014) were confirmed. The triangulation of the quantitative and qualitative data, helped to overcome the weaknesses of the statistical analysis, and also identified some of the intervening or controlling factors that may have been responsible for some students achieving higher test results than others. By interviewing individual students, critical qualitative information about their individual beliefs, perseverance, and behaviour could be explored.

The implications for teaching practice are that the evidence provided by this mixed methods study appeared to be sufficient, in terms of statistical and practical significance, to conclude that it is possible to make educational policy decisions that will benefit the acquisition of vocabulary by elementary Emirati learners of English. The ANOVA time x group interaction was not only significant at a high-level of statistical significance. The results also had practical significance, indicated by the moderate to large effect sizes. The researcher supported the arguments in the literature proposing that effect sizes are more important than p-values to evaluate the impact of educational interventions in experimental research designs (Hill & Thompson, 2004; Kotrlik & Williams, 2004; Lipsey, Puzio. Yun, Hebert, Steinka-Fry, et al., 2012). The moderate to large effect sizes (based on the criteria of Ferguson, 2009) obtained in this study implied that the effects of word cards and translation approach had practical significance. It can be concluded that the results of this mixed methods study have practical

implications, for example to assist the making policy of decisions in educational settings (Fraenkel & Wallen, 2011).

The results of this study provided evidence to support the policy of implementing decontextualized vocabulary study, using word cards and translation as an educational intervention, specifically for elementary level Emirati learners of English in UAE. However, the limited external and internal validity of the results (associated with the ecological fallacy, and the limitations of the statistical methodology) implied that decontextualised vocabulary study, using word cards and translation, may not necessarily promote a rapid gain in receptive vocabulary for all elementary Arab learners of English, at all times, and in all places.

Although these possible limitations are acknowledged, it is also important to highlight why this study may have wider implications for both Arab learners of English and wider, non-Arabic populations. Current western educational theory promotes a view that education should have critical thinking at its heart and that students should play a full role in a constructivist process, rather than solely relying on the teacher (Dahl, 2011). In the case of vocabulary learning, some have argued that no specific focus on the target language is required as students will acquire the vocabulary needed incidentally through reading (Krashen, 2013; Schmitt & Carter, 2000). These views assume that the students involved can already operate effectively in the language they will study in. This is clearly not the case for low level learners of English entering English medium institutions. An abrupt transition from the UAE state schooling system to the HCT western influenced higher education model may be particularly difficult for male Emirati students. Their educational experience prior to entering higher education at the time of this study in 2011 had been one of reliance on the teacher and remembering the knowledge given

(Hatherley-Greene, 2014). Therefore, it could be argued that Emirati language learners at CEFR A1-A2 levels, might be better served by the more traditional method of memorisation in their early efforts to acquire a working knowledge of the most frequent English vocabulary. If we accept that memorisation is an important component of vocabulary learning it would appear perfectly appropriate to utilise an approach both effective and familiar to the students. Thus, the continuity afforded by the word card approach when developing vocabulary size, may have the added advantage of lessening the shock of the transition to western educational methods as the students grapple with thinking critically in other areas.

However, it is also arguable that the benefits of memorisation in vocabulary learning are not restricted to Emirati students. As we have seen in the Vocabulary Learning Strategies section of the Literature Review, the key to the word card approach is that students are forced to retrieve meaning of vocabulary from memory and in doing so strengthen the links to the meaning on each occasion retrieval takes place. Studies on a range of populations are cited in the literature review including ones carried out in Malaysia, Taiwan and Turkey, give support to the effects of retrieval from language specialists (Ku & Ho, 2012; Ramachandran & Rahim, 2004; Royer, 1973). In addition, Phycologists and Psycholinguists strongly suggest that retrieval practice leads to retention of what is studied and develops memory (Kang, Lindsay, Mozer & Pashier, 2014; Roediger & Butler, 2011; Storm, Bjork & Storm. 2010). As there is no suggestion that the retrieval effect is peculiar to Emirati students, we can imply that the word card approach may well be effective with all populations of low level language learners and, indeed, with any learner wishing to focus on retaining vocabulary important to their studies.

Similarly, the planned recycling of vocabulary was also identified as a key element of vocabulary learning. The rationale for recycling is that as vocabulary is met on multiple occasions deeper knowledge of the word will be gained and recognition of the word will become automatic on sight. This in turn is likely to aid learners as they develop their reading skills. Again, there is no apparent reason why the benefits of recycling should be limited to Emirati learners.

Even when we consider the specific issues faced by Arab learners of English when developing breadth of vocabulary knowledge and reading skills, the solutions offered in this study appear to have implications for a wider population. For example, the literature highlights the difficulties faced by learners of English when their L1 has different orthography to English. The solution proposed and built into the word card treatment is for students to explicitly learn English spellings as they use the cards. This has clear applications for not only the wider Arab population but for all learners of English where their L1 has a different orthography.

This study also found that the students believed that the word card treatment was effective and that the participants preferred this strategy to others. If students' beliefs in a strategy reinforce their motivation to learn in this population it may also be the case in other populations.

Similarly, the surveys and interviews revealed (see tables 11& 30) that the participants perceived online learning to be effective; this was true for both the control and experimental groups. The key points appear to be speed of access to meaning and clarity. The online bi-lingual dictionaries were very popular with both groups and the online recycling activities, which formed part of the experimental treatment, were the approach that the experimental students

perceived as being most useful when learning vocabulary. This echoes the feelings of students in Taiwan and Hong Kong as described in the 'Online resources' section of the literature review. Therefore, it seems reasonable to suggest that approaches using well targeted online resources, as in this study, may have useful applications with other populations.

In summary, whilst the participants were all male Emiratis, as noted in the population and limitation sections of this study, research on developing vocabulary size using word cards has implications for all low level learners. The issues they faced are common issues for all Arab learners of English and are possibly the same issues faced by all learners of English with a different L1 orthography.

In addition, translation through word cards does appear to be an important ingredient in developing the vocabulary size (breadth) of English language learners and should be encouraged at low levels especially (Nation, 2008). How this might be achieved is discussed below.

The implications for teaching with elementary level Emirati learners include the possible adoption of a pre higher education online vocabulary course for students in the final two years of high school. This would have many advantages: it would not rely on the skill of the teacher; it would not encroach on classroom time; the time spent on the course could be monitored; and it would likely be an attractive manner of engaging in vocabulary learning for this age group. If it included the first 3,000 words of the British National Corpus (BNC) and the AWL, it could deliver a solid vocabulary platform to students preparing for their college foundation year. To this end two existing online programmes could be adapted and utilised. These programmes are 'Words to Know' (WTK) developed at Al Ain Women's College and a further programme

developed at Dubai Men's College (DMC). They are described in detail in the procedure section of the methodology chapter.

As discussed in the 'translation' section of the Literature Review chapter, translation is discouraged in the classrooms of the UAE. It is recommended that this view should be reconsidered. This does not mean that routine classroom discourse should be carried out in L1 but that judicious use of L1, where there is evidence of its effectiveness, as in the case of word cards, should be encouraged. A further implication here is that schoolteachers in the UAE may need training and support in deciding when the use of translation is desirable and when not.

The case for students reading expeditiously in higher education was made in the 'instruments' section of the Methodology chapter (Khalifa &Weir, 2009; Weir, Hawkey, Green, Unaldi & Devi, 2009). The seeding of this type of reading should also be encouraged at UAE schools through simple speed reading practice of texts at the 1,000 word level or below for elementary levels. These courses are readily and freely available (Millet, 2017). Writing similar texts in the context of stories about Emirati life and history would also be worth considering in order that the speed reading texts could be more relevant to Emirati students.

Recommendations for Future Research

The research in this study has focused on the effect of word cards in building vocabulary size. In essence word cards deal with the form-meaning link and promote automaticity of recall. The focus was firmly on building receptive vocabulary size breadth with only relatively minor attempts to develop depth of vocabulary knowledge through speed reading and edited reading texts. Future research could take this further by attempting to develop breadth and depth of vocabulary knowledge concurrently. Empirical research has shown that, in addition to receptive

vocabulary size (breadth), depth of vocabulary knowledge plays an important role in predicting reading ability as we have seen in the literature. That is, given two candidates with a similar vocabulary size, the one with greater depth of vocabulary knowledge is likely to have a higher level of reading comprehension (Qian, 1999). Schmitt (2014) suggests that with very low-level learners it is difficult to measure the gap between breadth and depth of knowledge. However, he goes on to argue that the gap becomes easily discernible at higher levels where breadth increases at a faster rate than depth. In view of this, it would be interesting to conduct research with learners with a vocabulary size of approximately 4,000 words. Schmitt and Schmitt (2014) found that there is a general lack of focus on any direct teaching of what they term 'mid-frequency' vocabulary, which they identify as the bands after the 3,000 level and up to and including the 9,000 word band. Therefore, the effect of a word card treatment on these levels, in tandem with exposure to this vocabulary in carefully edited texts and graded readers, might be a fruitful basis for research. The texts would need to be profiled and edited to ensure the maximum recycling in context of the target vocabulary. Research in this area could also investigate the impact of increases in breadth and depth knowledge of mid-frequency bands on reading proficiency. Laufer and Ravenhorst-Kalovski (2010) found that the minor increases in text coverage provided by knowledge of this mid-frequency vocabulary (0.8 percent increase in coverage from 5K to 6K), somewhat counter intuitively, contributed significantly large increases in reading test scores. This proved true for all mid-frequency bands. Daller and Xue's study (2009) highlighted the value of increases in vocabulary knowledge in an academic context. They found strong indications that knowledge of less frequent vocabulary improved the participants' academic writing performance and that pre university entry testing of lexical sophistication would be a

valuable addition to the screening of non-native students wishing to enter higher education in English medium institutions. The conclusion on the value of vocabulary testing in university entry procedures was supported in a later study (Daller & Phelan, 2013). The studies cited above underline the need for a greater research focus on the mid-frequency words in these often neglected bands.

Further interesting areas include targeted vocabulary work on academic vocabulary in specific disciplines. Evidence points to the AWL, which lists 570 academic words, providing 10.07 per cent coverage of research articles in the field of medicine (Chen & Ge, 2007). If, word lists were produced for individual disciplines, it seems plausible that the coverage would be considerably higher than that afforded by the more general AWL, with a relatively modest gain in specific vocabulary knowledge needed. Learning the initial meaning-form link would be covered by the word card treatment employed in the present study. As the vocabulary would be the highest frequency academic vocabulary specific to individual fields, the issues of ensuring constant recycling and exposure to the vocabulary in context would be met through reading the literature in the chosen field of study. Research into how much extra text coverage approximately 300 words would provide, the amount of study needed to reach automaticity of recall and what effect this would have on reading comprehension in the particular field, might all be profitable avenues for investigation.

In addition, it is recommended that the research project described in this study should be repeated in order to test for the confounding effects of controlling variables that may have threatened the internal validity of the results. These variables, which should be examined using quantitative and qualitative methods before, during, and after the experiment, could include: (a)

the amount of practice in English vocabulary and reading comprehensions that the students experienced at home, outside of the formal classroom sessions,; (b) the motivation levels of students (associated with their beliefs, perseverance, and behaviour); (d) the sensitivity of the teacher; and (e) the number of students in the class.

The recommended research assumes that the academic achievements of elementary Emirati learners of English with respect to learning vocabulary may not necessarily be directly related to the motivation of the students to respond to formal techniques designed only for teaching vocabulary. The motivation of the students could, for example, be related to their English language learning outside the classroom, the teacher's informal, intuitive, and spontaneous attitudes toward each individual student. Attempting to unravel these relationships would help researchers to improve the competencies of teachers of elementary Arab learners of English.

Finally, moving away from the UAE, current unpublished research on predictive analytics through vocabulary is underway. The vocabulary size of approximately 700 preparatory course students has been tested on entry to and completion of the Nazarbayev University (NU) foundation year programme for the past three years (Kinsella, 2018). Entry to the programme requires passing an IELTS test at a minimum of band 6 overall and the mean overall score is 6.7 on entry currently. However, each year, for various reasons, some students are not successful and do not enter the undergraduate programme. One reason for the research is to use the data at the entry point to identify students with vocabulary sizes of less than 7,000 words and then to monitor those students closely on assessments in the first half of semester 1. If their academic performance is weak, they are interviewed to ascertain the possible reasons and, if these reasons

are of an academic nature, further support is provided through the programmes' Academic Learning Centre (ALC). A second reason is to track the students through their academic careers at the university. NU is an English medium university where students spend five years, including the foundation year and four years as undergraduates. The tracking has the eventual additional aim of investigating the correlation between vocabulary size on entry, performance in reading assessments and the GPA on completion of a degree.

Final Remarks

These final remarks are based on the guidelines provided by Johns (2004, p. 18) who recommended that the final remarks of a reflexive researcher, after completing a research project, should involve the answering of reflexive questions, including 'To what extent did you act in tune with your personal values?' As a teacher of elementary level Arab learners of English, the researcher's personal values were based on his personal need to help rectify the situation, as described in the literature review, that elementary level Emirati learners of English have generally not been successful and may have developed a perception that what is demanded of them in terms of achieving high scores in English vocabulary and reading tests is too difficult for them (Davidson, Atkinson & Spring, 2011; Schmitt, 2008; Watts, 2011). From a professional point of view, the researcher considered himself duty bound to conduct research in a field that will help to improve existing knowledge and understanding of new strategies to improve vocabulary size and reading comprehension among elementary level Emirati learners of English.

Therefore, based on his personal values, the researcher is gratified that, despite its limitations, the main conclusions of this study were that (a) decontextualised vocabulary study,

using word cards and translation, appeared to contribute toward a more rapid gain in receptive vocabulary for elementary level Emirati learners of English than a similar teaching programme lacking this element; and that (b) the size of the receptive vocabulary appeared to correlate with reading comprehension scores. Furthermore, the philosophical stance of the researcher was vindicated, because the mixed methods research design, underpinned by constructive realism or pragmatism, generated quantitative data that was enriched and corroborated by qualitative data.

Finally, the researcher asserts reflexively that the journey he has travelled in order to collect, analyse, and interpret all the quantitative and qualitative data presented in this dissertation was long, difficult, and laborious. Nevertheless, this arduous and eventful journey equipped him with new knowledge and skills that will enable him to teach more effectively, and to conduct research more confidently, in a better and more meaningful way than before this journey started.

References

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics*. CA: Thousand Oaks: Sage

Afthanhorhan, W. (2013). A comparison of partial least squares structural equation modeling (PLS-SEM) and covariance based structural equation modeling (CB-SEM) for confirmatory factor analysis. *International Journal of Engineering and Innovative Technology, 2*, 198-205.

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: Wiley.

Alderson, P. (2004). Absence of evidence is not evidence of absence. *BMJ (Clinical Research Ed.), 328*(7438), 476-477. doi:10.1136/bmj.328.7438.476

Almekhlafi, A. G. (2006). The effect of computer assisted language learning (CALL) on United Arab Emirates English as a foreign language (EFL) school students' achievement and attitude. *Journal of Interactive Learning Research*, *17*(2), 121-142.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington DC: American Psychological Association.

Amineh, R. J., & Asi, H. D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages, 1*(1), 9-16.

Anthony, E. M. (1963). Approach, method and technique. *English Language Teaching, 17*, 63-67.

Arab-Moghaddam, N., & Senechal, M. (2001). Orthographic and phonological processing skills in reading and spelling in Persian/English bilinguals. *International Journal of Behavioral Development, 25*(2), 140-147.

Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of Management Education, 24*(1), 32-54.

Artino, A. R. (2012). Academic self-efficacy: From educational theory to instructional practice. *Perspectives on Medical Education, 1*(2), 76-85.

Babbie, E. R. (2010). *Basics of social research* (12th ed.). Belmont CA: Wadsworth.

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83-105). New York, NY: Psychology Press.

Bamford, J. (1993). Beyond grammar translation. In P. Wadden (Ed.), *A handbook for teaching English at Japanese colleges and universities* (pp. 63-71). Oxford: Oxford University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: WH Freeman and company.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*(1), 35-56.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language testing, 27*(1), 101-118.

Beglar, D. & Nation, P. (2007). A Vocabulary Size Test. *The Language Teacher, 31 (7),* 9-13.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24*(3-4), 323-335.

Biemiller, A., & Slonim, N. (2001). Estimating rootword vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498-520.

Birch, B. M. (2007). *English L2 reading: Getting to the bottom* (2nd ed.) Mahwah, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher, & A. Koriat (Eds.), *Attention and performance* (pp. 435-459). Cambridge, MA: MIT Press.

Bogdan, R. C., & Biklen, S. K. (2006). *Qualitative research in education: An introduction to theory and methods* (5th ed.) London: Allyn and Bacon.

Bowen, H. (2008). *Putting the spell in spelling.* (Unpublished MSc TESOL). Aston University,

Bowen, P., Rose, R., & Pilkington, A. (2017). Mixed methods-theory and practice. Sequential, explanatory approach. *International Journal of Quantitative and Qualitative Research Methods, 5*(2), 10-27.

Bowker, D., & Randerson, P. (2007). *Practical data analysis workbook for biosciences* (3rd ed.), Harlow: Pearson Education Ltd.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Brown, A. & Jaquith, P. (2011). The development and validation of an online rater training and marking system: promises and pitfalls. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (244-261). England: Palgrave Macmillan.

Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research, 15*(1), 83-97.

Brown, R. (1973). *A first language: The early stages.* Harvard University Press.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research, 1*(1), 8-22.

Burgmeier, A., & Zimmerman, C. B. (2007). *Inside reading 1:The academic word list in context.* New York: Oxford University Press.

Butler, S., Urrutia, K., Buenger, A., Gonzalez, N., Hunt, M., & Eisenhart, C. (2010). A review of the current research on vocabulary instruction. *National Reading Technical Assistance Center, RMC Research Corporation, 1.*

Cain, K., Lemmon, K., & Oakhill, J. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671-681.

Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Science, 3*(3), 106-116.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education, 42*(12), 1150-1152.

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review, 19*(1), 81-86.

Carter, R., & McCarthy, M. (1988). Developments in the teaching of vocabulary. In R. Carter, & M. McCarthy (Eds.), *Teaching English as a second or foreign language*. London: Longman.

Carver, R. (1993). The case against statistical significance testing revisited. *Journal of Experimental Education, 61*(4), 287-292.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Literacy Research, 26*(4), 413-437.

Cathain, A., & Murphy, E. (2010). Three techniques for integrating data in mixed methods studies. *British Medical Journal, 341.* doi: https://doi.org/10.1136/bmj.c4587

Chang, S. C. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English Language Teaching, 4*(2), 13-24.

Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26*(4), 502-514.

Chia, K. (1997). "Significant-itis"—an obsession with the P-value. *Scandinavian Journal of Work, Environment & Health,* 152-154.

Chien, C. W. (2015). Analysis the Effectiveness of Three Online Vocabulary Flashcard Websites on L2 Learners' Level of Lexical Knowledge. *English Language Teaching, 8*(5), 111-121.

Chingos, M., M., & Whitehurst, G. (2011). *Class size: What research says and what it means for state policy*. Washington, DC: Brookings Institute.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: Cambridge.

Cobb, T. (1995). Imported tests: Analysing the task. Al Ain, UAE: TESOL Arabia. Retrieved from http://www.lextutor.ca/cv/Al-Ain.html

Cobb, T. (2007). The old vocabulary, the new vocabulary, and the Arabic learner. In P. Davidson, C. Coombe, D. Lloyd & D. Palfreyman (Eds.), *Teaching and learning vocabulary in another language* (pp. 102-121). Dubai: TESOL Arabia.

Cobb, T. (2008). Commentary: Response to McQuillan and Krashen. *Language Learning & Technology, 12*(1), 109-114.

Cobb, T. (n.d.). Web VP Classic v.4 [computer program]. Retrieved from https://www.lextutor.ca/vp/eng/

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.

Conrad, N. J. (2008). From reading to spelling and spelling to reading: Transfer goes both ways. *Journal of Educational Psychology, 100*(4), 869-878.

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford: Oxford University Press.

Cook, V. J. (2004). *Accommodating broccoli in the cemetery: Or why can't anybody spell?* New York: Touchstone.

Coombe, C., & Davidson, P. (2014). Common Educational Proficiency Assessment (CEPA) in English. *Language Testing*, *31*(2), 269-276.

Corrigan, R. (2007). An experimental analysis of the affective dimensions of deep vocabulary knowledge in inferring the meaning of words in context. *Applied Linguistics, 28,* 211-240.

Creamer, E. G. (2018). *An introduction to fully integrated mixed methods research*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods* (4th ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., Plano Clark, & V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Crowder, M., & Hand, D. (1996). *Analysis of repeated measures. monographs on statistics and applied probability*. Boca Raton, FL: Chapman & Hall.

Cupchick, G. (2001). Constructivist realism: An ontology that encompasses positivist and constructivist approaches to the social sciences. *Forum: Qualitative Social Research, 2*(1), 29-39.

Dahl, M. (2010). *Failure to thrive in constructivism - a cross-cultural malady.* Transgressions – Cultural Studies and Education, Vol. 62. Rotterdam/Boston/Taipei: Sense Publishers.

Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review, 4*(1), 173-193.

Daller, M. H., & Xue, H. (2009). Vocabulary knowledge and academic success: A study of Chinese students in UK higher education. In *Vocabulary studies in first and second language acquisition* (pp. 179-193). London: Palgrave Macmillan.

Daniel, L. G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the Schools, 5*(2), 23-32.

Davidson, P., Atkinson, F., & Spring, J. (2011). The impact of explicitly teaching vocabulary on students' vocabulary learning. In *HCT, Foundations for the Future: Focus on Vocabulary* (*pp.* 27-41). Al Ain: HCT Press.

Davies, A. (2008). *Assessing academic English: Studies in language testing* (23rd ed.). Cambridge: Cambridge University Press.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom* Cambridge: Cambridge University Press.

Day, S. (2012). A reflexive lens: Exploring the dilemmas of qualitative methodology through the concept of reflexivity. *Qualitative Sociology Review, 8*(1), 60-85. Retrieved from https://www.qualitativesociologyreview.org/ENG/Volume21/QSR_8_1_Day.pd

Decarrico, J. (2001). Vocabulary learning and teaching. *Teaching English as a Second or Foreign Language, 3*, 285-299.

Denzin, N. K. (1997). Triangulation in educational research. In P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (2nd ed., pp. 318-322). London: Pergamon.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama, 263*(10), 1385-1389.

Douglas, N. (2009). *Reading Explorer 2*. Florence, KY, United States: Cengage Learning, Inc.

Duppenthaler, P. (2007). Vocabulary acquisition: The research and its pedagogical implications. *Tezukayama Gakuin University Research Journal, 42*, 1-14.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*, 55-63.

Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2012). *Management research – An Introduction.* London: Sage.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2*(1), 1-30.

Ehri, L. C. (2005). Learning to read words: Theory findings, and issues. *Scientific Studies of Reading, 9*(2), 167-188.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Ellis, N. (2008). Implicit and explicit knowledge about language. In J. Cenoz, & N. H. N. H. Hornberger (Eds.), *Encyclopedia of language and education* (pp. 1878-1890). USA: Springer.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891-904.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance--A critical literature review.* Chicago: Consortium on Chicago School Research.

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1), 19-42.

Ferguson, C. F. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538.

Ferrando, P. J., Anguiano-Carrasco, C., & Chico, E. (2011). The impact of acquiescence on forced-choice responses: A model-based analysis. *Psicologica: International Journal of Methodology and Experimental Psychology, 32*(1), 87-105.

Fetters, M. D., Curry, A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs-principles and practices. *Health Service Research, 48*(6), 2134-2156.

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks, CA: Sage.

Figueiredo F., Dalson B., Paranhos, R., Rocha, E. C., Batista, M., Silva , J. A., & Marino, J. G. (2013). When is statistical significance not significant? *Brazilian Political Science Review, 7*(1), 31-55.

Finlay, L. (2002). Negotiating the swamp: The opportunity and challenge of reflexivity in research practice. *Qualitative Research, 2*(2), 209-230.

Folse, K. (2004). Myths about teaching and learning second language vocabulary: What recent research says. *TESL Reporter, 37*(2), 1-13.

Folse, K. (2010). Is explicit vocabulary focus the reading teacher's job? *Reading in a Foreign Language, 22*(1), 139-160.

Fraenkel, J., & Wallen, N. (2000). *How to design and evaluate research in education*. New York: McGraw-Hill.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education* (7th ed.). New York: McGraw-Hill.

Fraser, C. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal, 91*(3), 372-394.

Fries, C. (1945). *Teaching and learning English as a foreign language* Ann Arbor: University of Michigan Press.

Frison, L., & Pocock, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine, 11*(13), 1685-1704.

Gairns, R., & Redman, S. (2008). *Oxford Word Skills: Learn and Practise English Vocabulary.* Oxford: Oxford University Press.

Gallo-Crail, R., & Zerwekh, R. (2002). Language learning and the Internet: Student strategies in vocabulary acquisition. In C. A. Spreen (Ed.), *New technologies and language learning: Cases in the less commonly taught languages* (pp. 55–79). Honolulu, HI: University of Hawai‘i, Second Language Teaching & Curriculum Center.

Gandara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs. *Policy Analysis for California Education, PACE (NJ1).*

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28*(2), 241-265.

Giddens, A. (1974). *Positivism and sociology.* London: Heinemann Educational Publishers.

Glass, G.V., P.D. Peckham, & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. *Reviews in Educational Research, 42*, 237-288.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*, 126-135.

Goodman, K. S. (1973). Psycholinguistic universals in the reading process. In F. Smith (Ed.), *Psycholinguistics and reading* (pp. 21-29). New York: Holt.

Grabe, W. (1991). Current developments in second language reading research. *Tesol Quarterly, 25*, 375-406.

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics, 24*(1), 44-69.

Grabe, W., & Jiang, X. (2014). Assessing reading. In J. A. Kunan (Ed.), *The companion to language assessment* (pp. 185-200). UK: Wiley- Blackwell.

Green, A., Ünaldi, A., & Weir, C. (2009). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing, 27*(2), 191-211.

Green, J., & Thorogood, N. (2009). *Qualitative methods for health research* (2nd ed.). Thousand Oaks, CA: Sage.

Gregg, K. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics, 5*(2), 79-100.

Grellet, F. (1981). *Developing reading skills.* Cambridge: Cambridge University Press.

Grotjahn, R., & Stemmer, B. (2002). C-tests and language processing. In J. A. Coleman, R. Grotjahn & U. & Raatz (Eds.), *University language learning and the C-test* (pp. 115-130) AKS-Verlag.

Guion, L. A. (2002). *Triangulation: Establishing the validity of qualitative studies* Department of Family, Youth and Community Sciences, University of Florida. Retrieved from https://sites.duke.edu/niou/files/2014/07/W13-Guion-2002-Triangulation-Establishing-the-Validity-of-Qualitative-Research.pdf

Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment Series, 1*, 1-6.

Hair, J., Anderson, R. E., Babin, B. J., Tatman, R. L., & Black, W. C. (2010). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: Sage.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*(1), 1-20.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle p-value generates irreproducible results. *Nature Methods, 12*(3), 179-185.

Hamada, Y. (2014). Japanese high school EFL learners' perceptions of strategies for preventing demotivation. *Asian EFL Journal, 75*(3), 4-20.

Handbooks for teachers (2005, 2007.2008). Cambridge, UK: University of Cambridge ESOL Examinations.

Harb, N., & El-Shaarawi, A. (2007). Factors affecting business students' performance: The case of students in United Arab Emirates. *Journal of Education for Business, 82*(5), 282-290.

Harrison, L., & Cushen, C. (2005). *Achieve IELTS 1: English for International Education.* Andover: Cengage Learning.

Harwell, M.R., Rubinstein, E.N., Hayes, W.S., & Olds, C.C. (1992). Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics, 17*, 315-339.

Hein, G. E. (2007). *Constructivist learning theory*. Lesley College Massachusetts, MA: Lesley College Press.

Hill, C. R., & Thompson, B. (2004). Computing and interpreting effect sizes. In J. C. Smart (Ed.), *Higher education handbook of theory and research* (XIX ed., pp. 175-196). Boston, MA: Klewer.

Hill, J., & Lewis, M. (2002). *LTP dictionary of selected collocations*. Boston, MA: Heinle & Heinle.

Hirsh, D. (1992). *The vocabulary demands and vocabulary learning opportunities in short novels.* (Unpublished MA). Victoria University of Wellington, New Zealand,

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689-689.

Horst, M. (2000). *Text encounters of the frequent kind: Learning L 2 vocabulary through reading.* (Unpublished PhD). University of Wales, Swansea,

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

House, E. R. (2008). Qualitative evaluation and social policy. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 623-640). Thousand Oaks, CA: Sage Publications.

Howatt, A. P. R. (1984). *A history of English language teaching*. Oxford: Oxford University Press.

Hu, H. C. M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with retention of word meanings inferred from context. *Canadian Modern Language Review, 68*(1), 54-77.

Hubbard, R., & Lindsay, R. M. (2008). Why p-values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology, 18*(1), 69-88.

Huberty, C. J., & Pike, C. J. (1999). On some history regarding statistical testing. *Advances in social science methodology, 5,* 1-22

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* New York: John Wiley.

Huck, S. W. (2009). *Statistical misconceptions*. New York, NY: Psychology Press.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-287). Cambridge: Cambridge University Press.

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2632-2640). Hoboken, New Jersey: Whiley-Blackwell. doi:10.1002/9781405198431.wbeal0530

Hurlbert, S. H., & Lombardi, C. M. (2009). Final collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian. *Annales Zoologici Fennici, 46*(5), 311-349.

Hurley, P. (1992). *A concise introduction to logic*. Belmont, CA: Wandsworth Publishing Company.

Hymes, D. H. (1972). On communicative competence. In J. B. Pride, & J. Holmes (Eds.), *Sociolinguistics: Selected readings.* (pp. 269-293). Harmondsworth: Penguin.

The IELTS handbook. (2007). Retrieved from https://vuonlen089.files.wordpress.com/2013/02/ielts-handbook-2007.pdf

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20.

Jamieson, S. (2004). Likert scales: How to (ab) use them. *Medical Education, 38*(12), 1217-1218.

Johns, C. (2000). *Becoming a reflective practitioner* (1st ed.). Oxford: Blackwell Science.

Johnson, P., & Duberley, J. (2003). Reflexivity in management research. *Journal of Management Studies, 40*(5), 1279-1303.

Johnson, R. B., & Christensen, L. B. (2004). *Educational research: Quantitative qualitative, and mixed approaches*. Boston, MA: Allyn and Bacon.

Jubran, S. (2012). Using multi sensory approach for teaching English skills and its effect on students' achievement at Jordanian schools. *European Scientific Journal, ESJ, 8*(22), 50-61.

Kabilan, M. K., Ahmad, N., & Abidin, M. J. Z. (2010). Facebook: An online environment for learning of English in institutions of higher education? *The Internet and Higher Education, 13*(4), 179-187.

Kang, S. H., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review, 21*(6), 1544-1550.

Karpicke, J., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250.

Kelly, L. G. (1969). *25 centuries of language teaching.* Rowley, MA: Newbury House Publishers.

Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System, 18*(2), 199-207.

Khalifa, H., & Weir, C. J. (2009). *Studies in language testing 29: Examining reading: Research and practice in assessing second language reading.* Cambridge & New York: Cambridge University Press.

Kinsella, L. (2018). *The use of vocabulary size in predictive analytics at Nazarbayev University.* Unpublished manuscript.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* NY: Cambridge University Press.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Research Journal, 21*(1), 1-7.

Krashen, S. (1982). *Principles and practice in second language acquisition* New York: Prentice-Hall.

Krashen, S. (1985). *The input hypothesis: Issues and implications* New York: Longman.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-464.

Krashen, S. (2003). *Explorations in language acquisition and use: The Taipei lectures*. Portsmouth, NH.: Heinemann.

Krashen, S. (2004). *The power of reading*. Portsmouth, NH.: Heinemann.

Krashen, S. (2013). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research, 1*(1), 27-43.

Kremmel, B. (2018). *Development and initial validation of a diagnostic computer-adaptive profiler of vocabulary knowledge* (Doctoral dissertation, University of Nottingham), Retrieved from eprints.nottingham.ac.uk. (49085).

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly, 13*(4), 377-392.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. Chicago: University of Chicago press.

Kuo, Y., & Ho, H. Y. (2012). Effects of word card strategy versus word list strategy on Taiwanese EFL junior high school students' vocabulary retention. *Electronic Journal of Foreign Language Teaching, 9*(1), 26–45.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.

Lange, R. (2012). *Testing local.* Paper presented at the 16th Annual Current Trends in English Language Testing Conference, Dubai, UAE.

Laufer, B. (1986). Possible changes in attitude towards vocabulary acquisition research. *IRAL: International Review of Applied Linguistics in Language Teaching, 24*(1), 73-79.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren, & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines.* (pp. 316-323). Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge: Cambridge University Press.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567-587.

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching, 42*(3), 341-354.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., Meara, P., & Nation, I. S. P. (2005). Ten best ideas for teaching vocabulary. *The Language Teacher, 29*(7), 11-14.

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *Eurosla Yearbook, 1*(1), 7-28.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal, 28*(1), 89-108.

Lee, M. J. (2014). Relative contribution of working memory capacity to L2 reading comprehension. *English Teaching, 69*(1), 119-146.

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove, UK: Language Teaching Publications.

Lewis, M. (1997). *Implementing the lexical approach.* Hove, UK: Language Teaching Publications.

Liao, P. (2006). EFL learners' beliefs about and strategy use of Translation in English learning. *RELC Journal, 37*(2), 191–215.

Linacre, J. M. (2019) Facets computer program for many-facet Rasch measurement, version 3.81.0 [computer software]. Beaverton, Oregon: Winsteps.com

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., & Cole, W. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. *United States: U.S.Dept of Education, National Center for Special Education Research, Institute of Education Sciences, NCSER, 2013-30.*

Liu, H. C., & Matthews, R. (2005). Vygotsky's philosophy: Constructivism and its critics examined. *International Education Journal, 6*(3), 386-399.

Lix L., Keselman J., Keselman H. (1996). Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance “F” Test. *Reviews in Educational Research, 66*, 579–619.

Lohman, D. F. (2006). *Understanding and predicting regression effects in the identification of academically gifted children.* Paper presented at AERA conference. Retrieved from http://faculty.education.uiowa.edu/dlohman/pdf/Understanding_and_predicting_regression.pdf

Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design: The case for the task. In G. Crookes, & S. Gass M. (Eds.), *Tasks in a pedagogical context: Integrating theory and practice*. Cleveland, UK: Multilingual Matters.

Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, *45*(1), 1-43.

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *AISHE Journal*, *9*(5), 3351-33514.

Mahmood, R. K. (2014). A pragmatic analysis of Yes/No questions in English with reference to press conferences. *Procedia-Social and Behavioral Sciences, 136*, 36-40.

Mares, C. (2003). Writing a coursebook. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 130-140). Cambridge: Cambridge University Press.

Mason, B., & Krashen, S. (2004). Is form-focused vocabulary instruction worthwhile? *RELC Journal*, *35*, 179-185.

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum: Qualitative Research, 11*(3)

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, *22*(1), 56-70.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3-11.

Maxwell, S. E., & Kelley, K. (2011). Ethics and sample size planning. In A. T. Panter, & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 159-183). New York, NY.: Routledge.

May, H., Erling, E., & Boruch, R. (2003). The ecological fallacy in comparative and international educational research: Discovering more from TIMSS through multilevel modeling. *CRESP Research Report.* Retrieved from https://repository.upenn.edu/gse_pubs/410

McGuinness, K. A. (2002). Of rowing boats, ocean liners and tests of the ANOVA homogeneity of variance assumption. *Austral Ecology, 27*(6), 681-688.

McKeown, M. G. (1993). Creating effective definitions for young word learners. *Reading Research Quarterly 28, 28*(1), 17-31.

McLaughlin, B. (1987). *Theories of second language learning*. London: Edward Arnold.

McLean, J. E., & Ernest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research*, *18*(4), 393-407.

Meara, P. (2010). *EFL vocabulary tests* (Second ed.). Swansea: Lognostics.

Meara, P., & Randall, M. (1988). How Arabs read Roman letters. *Reading in a Foreign Language, 4*(2), 133-145.

Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal, 5*(1), 4-18.

Merriam, S. (2014). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Millet, S. (2017). *Speed readings for ESL learners, 500 BNC. (ELI occasional publication no. 26).*Wellington, New Zealand: School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Mishima, T. (1967). An experiment comparing five modalities of conveying meaning for the teaching of foreign language vocabulary. *Dissertation Abstracts 27, 3030-3031A.*

Mitchell, R., & Myles, F. (1998). *Second language learning theories*. London: Edward Arnold.

Mittag, K. C., & Thompson, B. (2000). A national survey of AERA members perceptions of statistical significance tests and other statistical issues. *Educational Researcher, 29*, 14-20.

Mogashoa, T. (2014). Applicability of constructivist theory in qualitative educational research. *American International Journal of Contemporary Research, 4*, 51-59.

Mondal, N. K. (2012). English teachers' attitude towards grammar-translation method at secondary education: Bangladeshi context. *Researcher, 4*(2), 64-68.

Mondria, J., & Wit-De Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics, 12*(3), 249-267.

Moran-Elis, J., Alexander, V. D., & Cronin, A. (2006). Triangulation and integration: Processes, claims and implications. *Qualitative Research, 6*, 45-59.

Morgan, D. L. (2014). *Integrating qualitative and quantitative methods: A pragmatic approach*. Thousand Oaks, CA: Sage.

Na, L., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, *16*(1), 33-42.

Nagy, W. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description acquisition and pedagogy* (pp. 64-83). Cambridge, UK: Cambridge University Press.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown, & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, N. J.: ErIbaum.

Nagy, W. E., Herman, P. A., & Anderson, R. A. (1985). Leaming words from context. *Reading Research Quarterly, 20*, 233-253.

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology, 15*(6), 1044-1045.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: CUP.

Nation, I. S. P. (2003). Research and practice in vocabulary teaching: When will they meet? Paper presented at the *Proceedings of the 16th English Australia Education Conference,* Melbourne. Retrieved from http://pandora.nla.gov.au/pan/54720/20060120-

0000/www.englishaustralia.com.au/ea_conference03/proceedings/pdf/028F_Nation_Plenary.pdf

Nation, I. S. P. (2005). Teaching vocabulary. *Asian EFL Journal*, *7*(3), 47-54.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82.

Nation, I.S.P. (2013). *What should every EFL teacher know?* Compass Publishing Tokyo, Japan.

Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In M. H. Long, & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 543-559). Oxford: Blackwell Publishing Ltd. doi:https://doi.org/10.1002/9781444315783.ch28

Nation, P., & Ming-Tzu, K. W. (1999). Graded readers and vocabulary. *Reading in a foreign language*, *12*(2), 355-380.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. (pp. 6-19). Cambridge: Cambridge University Press.

Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle/Cengage.

Nesselroade, J. R., Stigler, S. M., & Bates, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, *88*(3), 622-637.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, *20 A*, 175-240; 263 - 294.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical Or Physical Character, Ser A*(231), 289-337.

Nix, T. W., & Barnette, J. J. (1998). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. *Research in the Schools, 5*(2), 3-14.

Norman, G. (2010). Likert scales, levels of measurement and the ‘‘laws’’ of statistics. *Advances in Health Science Education, 15*(5), 625-832. doi:10.1007/s10459-010-9222-y

Nuttall, C. (1982). *Teaching reading skills in a foreign language.* London: Heinemann Educational Books.

Nuzzo, R. (2014). Scientific method: Statistical errors. P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. *Nature, 506*(7487), 150-153.

O’Keefe, D. (2003). Colloquy: Should familywise alpha be adjusted? Against familywise alpha adjustment. *Human Communication Research, 29*(3), 431-447.

O’Sullivan, A. (2009). Reading in the gulf: Is it time for a re-focus? In D. Anderson (Ed,), *Cultivating real readers* (pp. 43-51). Al Ain, UAE: HCT Marifa.

O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS exit score gains in higher education. *International English Language Testing System (IELTS) Research Reports, 10*(3).

O'Loughlin, R. (2012). Tuning in to vocabulary frequency in coursebooks. *RELC Journal, 43*(2), 255-269.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In I. A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-384). Thousand Oaks, CA: Sage.

Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods, 15*(2), 199-228.

Ostyn, P., & Godin, P. (1985). RALEX: An alternative approach to language teaching. *Modern Language Journal*, *69*, 346-355.

Owens, R. E. (2001). *Language development: An introduction* (5th ed.). Needham Heights, MA: Allyn and Bacon.

Parana, J. (1998). *Lexical syllabus - A critical analysis.* (Unpublished Master's dissertation). Universidade Federal, Paraná.

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*(1), 09-29.

Parry, K. (1991). Building a vocabulary through academic reading. *Tesol Quarterly, 25*(4), 629-653.

Pathare, E. (2007). Eight aims for vocabulary materials design. In P. Davidson, C. Coombe, D. Lloyd & D. Palfreyman (Eds.), *Teaching and learning vocabulary in another language* (pp. 263-275). Dubai: TESOL Arabia.

Perfetti, C., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67-86). Washington, DC: American Psychological Association.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments? *British Medical Journal, 316*, 1236-1238.

Pfau, D. W. (1967). Effects of planned recreational reading programs. *The Reading Teacher, 21*(1), 34-39.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271-275.

Pogrow, S. (2017). 'The failure of the US education research establishment to identify effective practices: Beware effective practices policies'. *Education Policy Analysis Archives, 25*(5), 2-19.

Polak, J., & Krashen, S. (1988). Do we need to teach spelling? The relationship between spelling and voluntary reading among community college ESL students. *TESOL Quarterly, 22*(1), 141-146.

Prendergast, T. (2013). *The mastery of languages: Or the art of speaking foreign tongues idiomatically.* London: Forgotten Books.

Price, P., Jhangiani, R. and I-Chant, A. (2018). Experimental Research. In M. Crump, P. Price, R. Jhangiani, A. I-Chant & D. Leighton (EdS.), *Research Methods in Psychology.* Brooklyn College Edition. Available at: https://opentextbc.ca/researchmethods/chapter/experimental-design/ [Accessed 18 Mar. 2019].

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308.

Quinn, E., Nation, I. S. P. & Millet, S. (2018). Asian and Pacific speed reading for EFL learners. Retrieved from https://www.victoria.ac.nz/__data/assets/pdf_file/0007/1713175/Speed-reading-2018-charactcontrolled.pdf

Radnitzky, G., & Andersson, G. (1978). *Progress and rationality in science*. Dortrecht, Holland: Reidel Publishing Company.

Ramachandran, S. D., & Rahim, H. A. (2004). Meaning, recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *Relc*, *35*(2), 161-178.

Randall, M. (2009). Word recognition, psycholinguistics and teaching second language reading. In D. Anderson (Ed.), *Cultivating real readers* (pp. 13-23). UAE: HCT Press.

Raptis, H. (1997). Is second language reading vocabulary best learned by reading? *Canadian Modern Language Review*, *53*(3), 566-580.

Rasch D, Kubinger KD, Moder K: (2011) The two-sample t test: pretesting its assumptions does not pay. *Statistical Papers*, *52* (1), 219-231.

Rauber, A. S., & Gil, G. (2004). Feedback to grammar mistakes in EFL classes: A case study. *Revista Brasileira De Linguística Aplicada*, *4*(1), 277-289.

Richards, J. C. (1976). The role of vocabulary teaching. *Tesol Quarterly*, *10*, 77-89.

Richards, J. C. (2000). Series editor's preface. In N. Schmitt (Ed.), *Vocabulary in language teaching* (pp. xi-xii). Cambridge, England: Cambridge University Press.

Richards, J. C. (2001). The role of textbooks in a language program. *Guidelines*, *23*(2), 12-16.

Richards, J. C., & Rogers, T. S. (1986). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Richardson, P. M. (2004). Possible influences of Arabic-Islamic culture on the reflective practices proposed for an education degree at the Higher Colleges of Technology in the United Arab Emirates. *International Journal of Educational Development, 24*(4), 429-436.

Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *Views, 12*(2), 24-39.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning, 45*(5-6), 341-358.

Rivers, W. M. (1968). *Teaching foreign-language skills.* Chicago, University of Chicago Press.

Rivers, W. M. (1981). *Teaching foreign language skills.* Chicago: University of Chicago Press.

Rivers, W. M. (1983). *Speaking in many tongues: Essays in foreign-language teaching* (3rd ed.). Cambridge: Cambridge University Press.

Rochon, J., Gondan, M., & Meinhard, K. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology, 12*, 81.

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27.

Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (pp. 197-236). San Diego: Academic Press.

Rosenthal, R. (1992). Effect size estimation, significance testing, and the file-drawer problem. *Journal of Parapsychology, 56*, 57-58.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*(10), 1276.

Royer, J. M. (1973). Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports, 32*(1), 195-198.

Rumesy, L. (2013). *Shedding light on the predictive validity of English proficiency tests in predicting academic success.* (Unpublished Ph.D. Thesis). University of Newcastle upon Tyne, Newcastle, UK.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM approach*. Thousand Oaks, CA: Sage.

Ryan, A., & Meara., P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7*, 531-540.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System, 6*(2), 72-78.

Sawilowsky, S. (2005). Misconceptions leading to choosing the t-test over the Wilcoxon Mann–Whitney test for shift in location parameter. *Journal of Modern Applied Statistical Methods, 4* (2), 598–600.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration, 14*, 91-106.

Schanzenbach, D. W. (2012). Limitations of experiments in educational research. *Education Finance and Policy, 7*(2), 219-232.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1993a). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.

Schmidt, R. (1993b). Consciousness, learning, and interlanguage pragmatics. In G. Kasper, & S. Blum-Kulka (Eds.). *Interlanguage pragmatics*. New York: Oxford University Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language teaching and learning* (pp. 1-63). Honolulu: University of Hawai'I Press.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Schmieder, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *European Journal of Research Methods for the Behavioral and Social Sciences, 6*, 147-151.

Schmitt, N. (2002). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 827-840). Boston, MA: Springer.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Schmitt, N. (2010a). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.

Schmitt, N. (2010b). Key issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse & M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 28-40). Bristol, UK: Multilingual Matters.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913-951.

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *Tesol Journal, 9*(1), 4-9.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43.

Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition, and pedagogy.* Cambridge: Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55-88.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503.

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning, 10*(1), 4-33.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing, 24*(1), 99-128.

Shonfelder, W. (2011). CAQDAS and qualitative syllogism logic - NVivo 8 and MAXQDA 10 compared. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 12*(1). Retrieved from http://www.qualitative-esearch.net/index.php/fqs/article/viewFile/1514/3136

Shuttleworth, M., & Wilson, L. T. (2008). Research hypothesis. Retrieved from https://explorable.com/research-hypothesis

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666-681.

Sinclair, J. (1987). *Looking up: An account of the COBUILD project in lexical computing.* London & Glasgow: Collins ELT.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140-158). Harlow: Longman.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

Smeyers, P., & Depaepe, M. (2010). *Educational research - the ethics and aesthetics of statistics*. Berlin/Heidelberg, Germany: Springer Science and Business.

Smith, F. (1975). The role of prediction in reading. *Elementary English, 52*(3), 305-311.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology, 35*(1), 50-61.

Sorensen, R. (2017). Nothingness. Retrieved from https://plato.stanford.edu/archives/fall2017/entries/nothingness/

St John, W., & Johnson, P. (2000). The pros and cons of data analysis software for qualitative research. *Journal of Nursing Scholarship, 32*(4), 393-397.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139-152.

Stahl, S. A. (2005). Four problems with teaching word meanings. In E. H. . Hiebert, & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 95-114). Mahwah, NJ: Lawrence Erlbaum.

Stahl, S. A., & Kapinus, B. (2001). *Word power: What every educator needs to know about teaching vocabulary.* Washington, DC: National Education Association.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Stern, H. H. (1981). Communicative language teaching and learning: Toward a synthesis. In J. E. Alatis, H. B. Altman & P. M. Alatis (Eds.), *The second language classroom: Directions for the 1980's* (pp. 131-148). New York: Oxford University Press.

Sternberg, R. J. (1987). Most vocabulary is learnt from context. In M. G. McKeown, & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*(2), 244-253.

Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language, 33*(6), 551-571.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal, 25*(2), 34-50.

SurveyMonkey. [computer software]. San Mateo, California, USA. Retrieved from www.surveymonkey.com

Sweet, H. (1964). *The Practical Study of Language.* Oxford: Oxford University Press. (Original work published 1899)

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers of Human Neuroscience, 11*(390).

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 3-50). Thousand Oaks, CA: Sage.

The Oxford 3000. (n.d.). Retrieved from https://www.oxfordlearnersdictionaries.com/about/oxford3000

Thornbury, S. (1991). Metaphors we work by: EFL and its metaphors. *ELT Journal, 45*(3), 193-200.

Thornbury, S. (1998). The lexical approach: A journey without maps? *Modern English Teacher, 7*(4), 7-13.

Thornbury, S. (2018). An A-Z of ELT. Retrieved from https://scottthornbury.wordpress.com/2009/12/27/k-is-for-krashen/

Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge: Cambridge University Press.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*(2), 103-135.

Tunarosa, A., & Glynn, M. A. (2016). Strategies of integration in mixed methods research. *Organizational Research Methods, 20*(2), 224-242.

Twaddell, W. F. (1973). Vocabulary expansion in the TESOL classroom. *TESOL Quarterly, 7*, 61-78.

Twadell, F. (1972). Linguistics and the language teacher. In K. Croft (Ed.), *Readings on English as a second language* (pp. 268-276). Cambridge, Massachusetts: Winthrop.

Tyler-Smith, K. (2006). Early attrition among first time eLearners: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking eLearning programmes. *Journal of Online learning and Teaching*, *2*(2), 73-85.

UAE Ministry of Education. (2018). *The Emirates standardized test (EmSAT) to replace CEPA at universities*. Retrieved from https://www.moe.gov.ae/En/MediaCenter/News/pages/emsat.aspx

University of Cambridge ESOL Examinations. (2010). *Research notes* (Issue 40). Cambridge: University of Cambridge.

Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. *Educational and Psychological Measurement, 61*(2), 219-224.

Vickers, A.J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Medical Research Methodology, 5*(1), 35.

Wang, M., & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning, 55*(1), 71-98.

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition, 87*(2), 129-149.

Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World, 4*, 97-110.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108

Wasserstein, R., Schirm, A.L., & Lazar, N.A. (2019). Moving to a world beyond "p < 0.05". *The American Statistician, 73*(sup1), 1-9.

Watts, J. (2011). Measuring the effectiveness of a vocabulary programme. In *HCT, Foundations for the Future: Focus on Vocabulary*, (pp.15-26). Al Ain: HCT Press.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232-245.

Weeks, D. L. (2007). The regression effect as a neglected source of bias in nonrandomized intervention trials and systematic reviews of observational studies. *Evaluation and the Health Professions, 30*(3), 254-265.

Weir, C. J., Hawkey, R., Green, A., Unaldi, A., & Devi, S. (2009). The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a british university. *International English Language Testing System (IELTS) Research Reports, 9*, 97.

Weir, C. J., Vidaković, I., & Galaczi, E. D. (2013). *Measured Constructs: A History of Cambridge English examinations, 1913-2012.* Cambridge: Cambridge University Press.

West, M. (1926). *Learning to read a foreign language*. Bombay: Longmans Green and Company.

West, M. (1926/27). *The new method readers (new series) readers IA, IB, II, III,; companions IA, IB, II, III; supplementary readers I, II, III; Teacher's handbook*. Bombay & Calcutta: Longmans Green.

West, M. (1927). *The construction of reading material for teaching a foreign language*. Oxford: Oxford University Press.

West, M. (1953). The "reading approach" and "the new method system". *The Modern Language Journal, 22*, 220–222.

West, M. (1960). *Teaching English in difficult circumstances.* London: Longman.

Widdowson, H. G. (1987). Aspects of syllabus design. In M. Tickoo (Ed.), *Language Syllabuses: State of the Art.* Singapore: RELC.

Widdowson, H. G. (1978). *Teaching language as communication.* Oxford: Oxford University Press.

Wilkins, D. (1972). *Linguistics in language teaching.* London: Edward Arnold.

Willis, D. (1990). *The lexical syllabus: A new approach to language teaching.* London: Collins ELT.

Willis, D., & Willis, J. (1988). *Collins COBUILD English course.* London: Collins.

Wong, K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin, 24*(1), 1-32.

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education, 48*(2), 312-325.

Yin, R. K. (2014). *Case study research, design and methods* (5th ed.). Newbury Park, CA: Sage.

Yip, F. W., & Kwan, A. C. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43*(3), 233-249.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 57*(4), 541-572.

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.

Zimmerman, D. W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology, 127*(4), 354-364.

Appendix A

Questionnaire Administered to Control Group (Survey Monkey)

This questionnaire is designed to help me understand how you feel about different ways of learning English vocabulary. How you answer is very important and may help to develop our approach to vocabulary teaching and learning. Please take your time and think carefully about your answers.

Please grade the following ways of learning vocabulary from 1= not useful at all to 5 = extremely useful according to your experience on this course.

1. Learning vocabulary through graded reading books. For example: guessing a word you do not understand from the context of the story.
A not useful at all B slightly useful C useful D very useful E extremely useful

2. Learning vocabulary from exercises before you read a text. For example: putting words into sentences or matching words with definitions.
A not useful at all B slightly useful C useful D very useful E extremely useful

3. Learning vocabulary through communication. For example: talking to teachers or other students at college.
A not useful at all B slightly useful C useful D very useful E extremely useful

4. Learning vocabulary through online vocabulary sites .

A not useful at all B slightly useful C useful D very useful E extremely useful

5. Learning vocabulary through vocabulary books. For example: the Oxford Word Skills book.
A not useful at all B slightly useful C useful D very useful E extremely useful

6. Learning vocabulary through making your own lists. For example: vocabulary record sheets.
A not useful at all B slightly useful C useful D very useful E extremely useful

7. Learning vocabulary through using an English only dictionary. For example: noting the word and finding the definition.
A not useful at all B slightly useful C useful D very useful E extremely useful

8 Learning vocabulary through using an English/Arabic dictionary. For example: noting the word and finding the translation in the Oxford Word power English/Arabic dictionary.
A not useful at all B slightly useful C useful D very useful E extremely useful

Reading

This part of the questionnaire is designed to find out what you think are your main problems when you read in English.

1 My reading speed is a problem. For example: I can't read fast enough in exams.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

2 Unknown vocabulary is a problem.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

3 English text layout is different and makes it difficult to read. For example: Arabic is read from the right to the left but English is left to right.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

4 I know little about the topics in the reading texts and this makes it difficult.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

5 The grammar of English makes reading difficult.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

Appendix B

Questionnaire Administered to Experimental Group (Survey Monkey)

This questionnaire is designed to help me understand how you feel about different ways of learning English vocabulary. How you answer is very important and may help to develop our approach to vocabulary teaching and learning. Please take your time and think carefully about your answers.
Please grade the following ways of learning vocabulary from 1= not useful at all to 5 = extremely useful according to your experience on this course.

1 **Learning vocabulary through graded reading books. For example: guessing a word you do not understand from the context of the story.**
A not useful at all B slightly useful C useful D very useful E extremely useful

2 **Learning vocabulary from word cards with English on one side and the Arabic translation on the other.**
A not useful at all B slightly useful C useful D very useful E extremely useful

3 **Learning vocabulary from exercises before you read a text. For example: putting words into sentences or matching words with definitions**.
A not useful at all B slightly useful C useful D very useful E extremely useful

4 **Learning vocabulary through communication. For example: talking to teachers or other students at college**.
A not useful at all B slightly useful C useful D very useful E extremely useful

5 Learning vocabulary through online vocabulary sites .
A not useful at all B slightly useful C useful D very useful E extremely useful

6 Learning vocabulary through vocabulary books.
A not useful at all B slightly useful C useful D very useful E extremely useful

7 Learning vocabulary through making your own lists. For example: vocabulary record sheets.
A not useful at all B slightly useful C useful D very useful E extremely useful

8 Learning vocabulary through using an English only dictionary. For example: noting the word and finding the definition.
A not useful at all B slightly useful C useful D very useful E extremely useful

9 Learning vocabulary through using an English/Arabic dictionary. For example: noting the word and finding the translation in the Oxford Word power English/Arabic dictionary.
A not useful at all B slightly useful C useful D very useful E extremely useful

Reading (Survey Monkey)

This part of the questionnaire is designed to find out what you think are your main problems when you read in English.

1 My reading speed is a problem. For example: I can't read fast enough in exams.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree\

2 Unknown vocabulary is a problem.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

3 English text layout is different and makes it difficult to read. For example: Arabic is read from the right to the left but English is left to right.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

4 I know little about the topics in the reading texts and this makes it difficult.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

5 The grammar of English makes reading difficult.
A Strongly disagree B disagree C neither agree nor disagree D agree E strongly agree

Appendix C

Work Plans

Vocabulary	**Experimental: What**	**When and level**	**Control: What**	**When and level**
	Word Cards	Throughout the study	Oxford Word Skills Basic Units	Low-level Semester 1
			Oxford Word Skills Inter. Units 1-34	Low-level Semester 2 High-level Semester 1
			Oxford Word Skills Inter. Units 26-58	High-level Semester 2
	Words to know: online practice	Throughout the study	Inside Reading 1 (High AWL) Units 1-10	High-level Semester 2
	DMC: online practice	Throughout the study	Graded Readers (All- above VS) 10 books	Low-level Semester 1
			Reading Explorer 1 (low-level)	Semester 1
Reading	Graded Readers (at or below current VS)	Throughout the study	Graded Readers (above VS)	Throughout the study for both control levels
	Speed Reading (at or below current VS)	Semester 2 for both experimental levels	Reading Explorer 1 (low)	Low-level Semester 1
			Reading Explorer 2 (low & high)	Semester 2 low Semester 1 high
	Graded intensive reading (at or below current VS)	Throughout the study	Achieve IELTS 1 (High-level)	Semester 2
			Inside Reading 1 (High-level AWL)	Semester 2
Writing	Great Sentences for Great Paragraphs All units (1-8) (low)	Semester 1	Great Sentences for Great Paragraphs All units (1-8) (low)	Semester 1
	Effective Academic Writing 2 (low)	Semester 2	Effective Academic Writing 2 (low)	Semester 2
	Achieve IELTS 1 (high)	Semester 2	Achieve IELTS 1 (High-level)	Semester 2
Speaking	Tune In 2 (low)	Semesters 1&2	Tune In 2 (Low-level)	Semesters 1&2
	Achieve IELTS 1 (high)	Semester 2	Achieve IELTS 1 (High-level)	Semester 2
Listening	Expanding Tactics for Listening (low)	Semester 1	Expanding Tactics for Listening	Semester 1

	Tune In 2 (low and high)	Semester 2 (low) Semester 1 (high)	Tune In 2 (low)	Semester 2 (low) Semester 1 (high)
	Achieve IELTS 1 (high)	Semester 2	Achieve IELTS 1 (high)	Semester 2
Grammar	Center Stage 2 Grammar to Communicate. (low)	Semester 1	Center Stage 2 Grammar to Communicate.	Semester1
	Oxford Living Grammar - Pre-Intermediate (low)	Semester 1 (high-level) Semester 2 (low-level)	Oxford Living Grammar - Pre Intermediate (low)	Semester 1 (high-level) Semester 2 (low-level)
	Oxford Living Grammar Intermediate (high)	Semester 2	Oxford Living Grammar Intermediate (high)	Semester 2

Appendix D

Consent

Research Consent Form

'To what extent and in what ways does an intensive programme of targeted intentional vocabulary instruction impact on the development of reading proficiency?'

Dear student,
This study is being conducted by: Laurence Kinsella, one of your teachers, who is also studying at CRELLA, University of Bedfordshire.
I would like to invite you to participate in a research project. The research will investigate two types of vocabulary teaching and learning. I am doing this as part of a Doctoral degree at the University of Bedfordshire in the UK.

Procedure

We will do the following:

- All groups will focus on improving their vocabulary.
- I will measure your vocabulary size at the beginning, during and at the end of your course, and some aspects of your reading and other learning.
- I will compare the results to discover if one approach gives better results than the other.
- During and at the end of the course I will ask you to complete a short online questionnaire and I may also ask you to do a short interview, about how you felt about the way you learnt vocabulary and other areas of your learning experience.

Risks and Benefits of Being in the Study:

The study possesses no risks. All of the students involved will be taught vocabulary through materials and methods designed by experts in English language teaching and the research will not affect your normal studies in any way.
The possible benefits include:

Your teachers will gain information on the most efficient ways for you to learn vocabulary. This may help you in your future studies at the HCT. Your vocabulary and reading might get better.

Confidentiality:

The records of this study will be kept private. In any sort of report that might be published, I will not include any information that will make it possible to identify you as a participant. Research records will be kept in a locked file and fingerprint access computer; only the researcher and supervisory panel will have access to the records.

Voluntary Nature of the Study:
Your decision whether or not to participate will not affect your current or future relations with the researcher, your teachers or the HCT. If you decide to participate, you are free to withdraw at any time without affecting those relationships. You are also free to withhold any information you feel unhappy about sharing.

The researcher conducting this study is Laurence Kinsella. The researcher's supervisor is Stephen Bax. You may ask any questions you have now. If you have questions later, I will be available throughout your course and will provide post course contact details.

You will be given a copy of this form to keep for your records. I will tell you the results of the study if you wish to know them.
Laurence Kinsella, Al Ain Men's College, P O Box 17155, Al Ain
Phone: 03 7095647 Email: lkinsella@hct.ac.ae Room: G14

At the end of this, I hope to use the results to improve your vocabulary learning and your reading. I hope you can help me.
Thank you and regards,

Laurence

Statement of Consent:
I have read the above information. I have had the opportunity to ask questions and receive answers. I consent to participate in the study.

Printed Name of Participant: _____

Signature: _____ Date: _____

Signature of the researcher: _____ Date: _____

Appendix E: Ethical Documentation

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Annex to RS1 form)

SECTION A To be completed by the candidate

Registration No: 0924382

Candidate: Laurence Kinsella

Research Institute: Bedfordshire Business School

Research Topic: To what extent and in what ways does targeted intensive vocabulary instruction impact on the development of reading proficiency"

External Funding: No

The candidate is required to summarise in the box below the ethical issues involved in the research proposal and how they will be addressed. In any proposal involving human participants the following should be provided:

- clear explanation of how informed consent will be obtained,
- how will confidentiality and anonymity be observed,
- how will the nature of the research, its purpose and the means of dissemination of the outcomes be communicated to participants,
- how personal data will be stored and secured
- if participants are being placed under any form of stress (physical or mental) identify what steps are being taken to minimise risk

If protocols are being used that have already received University Research Ethics Committee (UREC) ethical approval then please specify. Roles of any collaborating institutions should be clearly identified. Reference should be made to the appropriate professional body code of practice.

Informed consent

The students involved in this study are adult entry level foundation students at Al Ain Men's College, UAE.

The participants will take part in an orientation week when the nature of the new approaches to foundation teaching will be explained. Participants will be made aware that data collected will be used to compare two approaches to vocabulary learning/teaching and how these affect reading skills development.

Participants will be provided with copies of a consent form in English and Arabic, which they will sign and return if they consent. An Arabic speaking member of staff will also be present to clarify any questions arising.

Confidentiality

All participants will be assigned a group and student number to ensure individual identities remain secure.

The records of this study will be kept private. In any sort of report that might be published, I will not include any information that will make it possible to identify a subject.

Data Storage

Research records will be kept in a locked file and fingerprint access computer; only the researcher and supervisor will have access to the records.

Permission for access

The Chair of English and the Director of Al Ain Men's College have sanctioned the study. I have written confirmation, in the form of an email, from the Chair of English.

Answer the following question by deleting as appropriate:

1. Does the study involve vulnerable participants or those unable to give informed consent (e.g. children, people with learning disabilities, your own students)?
They are my students but since they are adults are all able to give consent - college not school students

2. Will the study require permission of a gatekeeper for access to participants (e.g. schools, self-help groups, residential homes)?
Yes, this has been obtained.

3. Will it be necessary for participants to be involved without consent (e.g. covert observation in non-public places)?
No

4. Will the study involve sensitive topics (e.g. sexual activity, substance abuse)?
No

5. Will blood or tissue samples be taken from participants?
No

6. Will the research involve intrusive interventions (e.g. drugs, hypnosis, physical exercise)? **No**

7. Will financial or other inducements be offered to participants (except reasonable expenses)? **No**

8. Will the research investigate any aspect of illegal activity?
No

9. Will participants be stressed beyond what is normal for them?
No

10. Will the study involve participants from the NHS (e.g. patients or staff)?
No

If you have answered yes to any of the above questions or if you consider that there are other significant ethical issues then details should be included in your summary above. If you have answered yes to Question 1 then a clear justification for the importance of the research must be provided.

*Please note if the answer to Question 10 is yes then the proposal should be submitted through **NHS research ethics approval procedures** to the appropriate **COREC.** The UREC should be informed of the outcome.

Checklist of documents which should be included:

- Project proposal (with details of methodology) & source of funding
- Documentation seeking informed consent (if appropriate)
- Information sheet for participants (if appropriate)
- Questionnaire (if appropriate)

Signature of Applicant: Laurence Kinsella Date: 09/09/10

Signature of Director of Studies: Date: 19/9/10

This form together with a copy of the research proposal should be submitted to the Research Institute Director for consideration by the Research Institute Ethics Committee/Panel

Note you cannot commence collection of research data until this form has been approved

SECTION B To be completed by the Research Institute Ethics Committee:

Comments: Approved

Signature Chair of Research Institute Ethics Committee:

Date: 12/10/10

This form should then be filed with the RS1 form

If in the judgement of the committee there are significant ethical issues for which there is not agreed practice then further ethical consideration is required before approval can be given and the proposal with the committees comments should be forwarded to the secretary of the UREC for consideration.

Signature Chair of Research Institute Ethics Committee: Date:

This form together with the recommendation and a copy of the research proposal should then be submitted to the University Research Ethics Committee

Appendix F

Data Collection table

Data was collected over the academic year 2010/11. This consisted of two 19 week semesters, semester 1 and 2.

When	Pre College	Semester 1 Weeks 2&3	Semester 1 Week 12	Semester 1 Week 16	Semester 2 Week 2	Semester 2 Week 4	Semester 2 Week 9	Semester 2 Week 13
Instrument								
CEPA	X							
VST		X		X		X	X	
VLT		X		X				
AWL Test					X			
IELTS		X	X		X		X	
PET			X					
Questionnaire								
Interviews								
Attendance Records	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study
When	Semester 2 Week 15	Semester 2 Week 16	Semester 2 Week 17	Semester 2 Week 18				
Instrument								
CEPA								
VST		X						
VLT	X							
AWL Test		X						
IELTS		X						
PET			X					
Questionnaire			X					
Interviews				X				
Attendance Records	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	Throughout study	